

明細書

トランSPORT層中継方法及びトランSPORT層中継装置並びにプログラム技術分野

[0001] 本発明は、パケット交換による通信システムに関し、特に、輻輳制御機能を有するプロトコル層のコネクションを中継する技術に関する。

背景技術

[0002] OSI参照モデルのネットワークプロトコルでは、データ通信ネットワークの機能を7つの階層に大別しており、下位の階層から順に物理層、データリンク層、ネットワーク層、トランSPORT層、セッション層、プレゼンテーション層、アプリケーション層がある。パケット網において、端末上のアプリケーションが、複数のパケットにまたがるデータを通信する場合、アプリケーション間で仮想的なコネクションが張られ、そのコネクション上にアプリケーションデータが送られる。このコネクションは、一般に、トランSPORTより上位層のプロトコルを用いて張られる。また、このコネクションを張る層のプロトコルの中には、受信側の受信処理能力を超えないように送信側のレートを制御するフロー制御、および、途中のネットワークの輻輳を起こさないように、送信側のレートを制御する輻輳制御の両方の機能(両方を合わせてレート制御と呼ぶ)を持つものが存在する。また、レート制御は、受信側から送信側へフィードバック情報(正しく受信したデータのシーケンス番号の情報や、受信側の受信バッファの残り量など)を送り返して、送信側の送信レートを制御するフィードバック制御を行うものが多い。一般に、この輻輳制御機能はトランSPORT層で実現される事が多いため、以下では、上記の輻輳制御に関わるプロトコルが送受信アプリケーションの間で張るコネクションを、トランSPORT層コネクションと呼ぶことにする。現在インターネットで広く用いられているTCP(Transmission Control Protocol、Transport Control Protocolとも呼ばれる)も、トランSPORT層プロトコルであり、輻輳制御機能を持っている。

[0003] トランSPORT層プロトコルの輻輳制御機能は、それぞれのコネクションが次のように互いに独立にレート制御を行う。即ち、複数のトランSPORT層コネクションが、あるリンク上に同時に存在し、リンクの帯域を競合して取り合う場合、それらのコネクションの

送信側の輻輳制御を含めたトランSPORT層プロトコルの処理を行う部分(トランSPORT層コネクション終端部と呼ぶ)が、各コネクション毎に独立にレート制御を行うのである。TCPでは、受信端末からのACKパケットを元に、パケットロスや、遅延の増加を検知して、リンクの帯域を全コネクションで使い切って、かつ、コネクション間で取り合う帯域がほぼ同じになるような制御が行われる。このため、特定のコネクションに高い帯域を割り当てたり、コネクション間で帯域を融通しあって、効率よい通信を行う事が出来なかつた。また、トランSPORT層の輻輳制御機能は、コネクション開設後に、ネットワークの輻輳状況を学習する事によって適切なレートでの送信を行つてゐる。このため、コネクション開設直後は通信効率がよくなかった。

- [0004] これらの問題に対処するため以下のような従来技術が存在する。
- [0005] 第1の従来技術は、ルータのパケットスケジューリングによる方法で、特定のコネクションに属するパケットを識別し、特定のフローに高い帯域を割り当てたり、コネクション間で優先度をつけたりする方式である。
- [0006] 第2の従来技術は、Congestion Managerと呼ばれる方式で、ある端末の組(例えば端末Aと端末B)の間に張られた複数のトランSPORT層コネクションについて、それぞれのコネクションの送信レートについて独立に輻輳制御するのではなく、全体の送信レートを1本のTCPと同じ輻輳制御で得られるレートにし、それぞれのコネクションにその1本の送信帯域を割り当てる方式である。
- [0007] 第3の従来技術は、複数の端末のTCPを一旦終端し、1つのTCPに多重して送り、受信端末の近くで再びもとのTCPへと分離して中継する方式である(例えば、特開2002-185488号公報参照)。
- [0008] 従来方式の第1の問題点は、第1の従来技術では、送受信端末の通信パス上の全てのルータにおいて、パケットスケジューラでの優先度制御、固定帯域割り当て等の制御を行わなければならず、多数のルータを置き換える必要が生じ、多数の装置コスト、および、装置の入れ替えにかかるコストが発生することにある。
- [0009] 従来方式の第2の問題点は、第2の従来技術、および、第3の従来技術では、着目しているコネクション全体のスループットの制御を、既存の1本のTCPの輻輳制御を用いて行つてゐるため、着目しているトラヒック全体の実効レート(実際に受信側に届い

たデータのスループット)をある所望の値にする事、および、着目しているトラヒック以外のトラヒックとの間の優先付けが出来ない事にある。

- [0010] 従来方式の第3の問題点は、第2の従来技術では、同一の端末の組の間に張られた複数のトランスポート層コネクションについて輻輳制御の協調は出来るが、異なる端末の組の間に張られたトランスポート層コネクションについての輻輳制御の協調は出来ない事にある。
- [0011] 従来方式の第4の問題点は、第3の従来技術では、端末の初期の送信レートがTCPのスロースタートとなり、レートが適切な値となるのに時間がかかるという事にある。

発明の開示

- [0012] 本発明の第1の目的は、複数の端末間のあるトランスポート層コネクション集合に対して、所望する実効レートを確保し、そのコネクション集合の中で、その帯域を再分配することを、低成本で実現する事にある。
- [0013] 本発明の第2の目的は、複数の端末間のあるトランスポート層コネクション集合に対して、他のトランスポート層コネクションに対して優先度付けした実効レートを確保し、そのコネクション集合の中で、その帯域を再分配することを、低成本で実現する事にある。
- [0014] 本発明の第3の目的は、複数の端末間のあるトランスポート層コネクション集合に対し、ある実効レートを、そのコネクション集合の中で、コネクションを流れるアプリケーションの情報に依存して再配分する事にある。
- [0015] 本発明の第4の目的は、複数の端末間のあるトランスポート層コネクション集合に対して、所望する実効レートの確保を、低成本かつ高精度で実現する事にある。
- [0016] 本発明の第5の目的は、複数の端末間のあるトランスポート層コネクション集合に対して、他のトランスポート層コネクションに対して優先度付けした実効レートの確保を、高精度で実現する事にある。
- [0017] 本発明の第6の目的は、端末へ初期送信レートを伝えることで、端末の輻輳制御の効率を向上させる事にある。
- [0018] 上記の第1、2、3の目的を達成するために、本発明の第1のトランスポート層中継装置は、端末からのトランスポート層コネクションを終端する複数の端末側コネクション

終端部を持ち、それを対向するトランスポート層中継装置へ中継するためのトランスポート層コネクションを(中継コネクション)終端する複数の装置間コネクション終端部を持ち、各端末側コネクション終端部と、各装置間コネクション終端部の間で、アプリケーション情報を読みとりながら、データを中継する、アプリケーション情報解析部を持つ。さらに、装置間コネクション終端部の総送信レートと、各装置間コネクション終端部の送信レートを決定する、送信レート制御部と、複数の端末側コネクション終端部、装置間コネクション終端部、アプリケーション情報解析部のどれが使われているかを管理する中継制御部を持つ。

- [0019] 上記の第4、5の目的を達成するために、本発明の第2のトランスポート層中継装置は、第1のトランスポート層中継装置が持つ構成に加え、対向する装置間のネットワーク状況を推定し、送信レート制御部へ通知するネットワーク状況推定部を持つ。
- [0020] 上記の第1、2、3、の目的を達成するために、本発明の第3のトランスポート層中継装置は、端末からのトランスポート層コネクションを終端する複数の端末側コネクション終端部を持ち、それを対向するトランスポート層中継装置へ中継するトランスポート層コネクション(中継コネクション)を終端する1つの装置間コネクション終端部を持つ。また、各端末側コネクション終端部からのデータを集約して1つの装置間トランスポート層終端部へ受け渡すMUX-DEMUX部を持つ。また、各端末側コネクション終端部と、MUX-DEMUX部の間で、アプリケーション情報を読みとりながら、データを中継する、アプリケーション情報解析部を持つ。また、装置間コネクション終端部の総送信レートと、MUX-DEMUX部がどのようなレートの割合で各端末側コネクション終端部からのデータを集約するかを指示する、送信レート制御部と、複数の端末側コネクション終端部、アプリケーション情報解析部のどれが使われているかを管理する中継制御部を持つ。
- [0021] 上記の第4、5の目的を達成するために、本発明の第4のトランスポート層中継装置は、第3のトランスポート層中継装置が持つ構成に加え、対向する装置間のネットワーク状況を推定し、送信レート部へ通知するネットワーク状況推定部を持つ。
- [0022] 上記の第6の目的を達成するために、本発明の第5のトランスポート層中継装置は、本発明の第1のトランスポート層中継装置のの中で、端末からのコネクション開設要求

時に、端末に通知すべき最適な初期送信レートを送信レート制御部に問い合わせせる機能が加わった中継制御部と、その要求に応じて最適な初期送信レートを決定する機能が加わった送信レート制御部手段と、端末に初期送信レートを通知する機能が加わった端末側コネクション終端部を持つ。

- [0023] 上記の第6の目的を達成するために、本発明の第6のトランスポート層中継装置は、本発明の第2のトランスポート層中継装置の構成の中で、端末からのコネクション開設要求時に、端末に通知すべき最適な初期送信レートを送信レート制御部に問い合わせせる機能が加わった中継制御部と、その要求に応じて最適な初期送信レートを決定する機能が加わった送信レート制御部と、端末に初期送信レートを通知する機能が加わった端末側コネクション終端部を持つ。
- [0024] 上記の第6の目的を達成するために、本発明の第7のトランスポート層中継装置は、本発明の第3のトランスポート層中継装置の構成の中で、端末からのコネクション開設要求時に、端末に通知すべき最適な初期送信レートを送信レート制御部に問い合わせせる機能が加わった中継制御部と、その要求に応じて最適な初期送信レートを決定する機能が加わった送信レート制御部と、端末に初期送信レートを通知する機能が加わった端末側コネクション終端部を持つ。
- [0025] 上記の第6の目的を達成するために、本発明の第8のトランスポート層中継装置は、本発明の第4のトランスポート層中継装置の構成の中で、端末からのコネクション開設要求時に、端末に通知すべき最適な初期送信レートを送信レート制御部に問い合わせせる機能が加わった中継制御部と、その要求に応じて最適な初期送信レートを決定する機能が加わった送信レート制御部と、端末に初期送信レートを通知する機能が加わった端末側コネクション終端部を持つ。
- [0026] 本発明には以下の効果がある。
- [0027] 本発明の第1の効果は、複数の端末間のあるトランスポート層コネクション集合に対して、所望する実効レートを確保し、そのコネクション集合の中で、その帯域を再分配することを、低コストで実現できる事にある。
- [0028] 本発明の第2の効果は、複数の端末間のあるトランスポート層コネクション集合に対して、他のトランスポート層コネクションに対して優先度付けした実効レートを確保し、

そのコネクション集合の中で、その帯域を再分配することを、低コストで実現できる事にある。

- [0029] 本発明の第3の効果は、複数の端末間のあるトランスポート層コネクション集合に対し、ある実効レートを、そのコネクション集合の中で、コネクションを流れるアプリケーションの情報に依存して再配分できる事にある。
- [0030] 本発明の第4の効果は、複数の端末間のあるトランスポート層コネクション集合に対して、所望する実効レートの確保を、低コストかつ高精度で実現できる事にある。
- [0031] 本発明の第5の効果は、複数の端末間のあるトランスポート層コネクション集合に対して、他のトランスポート層コネクションに対して優先度付けした実効レートの確保を、高精度で実現できる事にある。
- [0032] 本発明の第6の効果は、端末へ初期送信レートを伝えることで、端末の輻輳制御の効率を向上できる事にある。

図面の簡単な説明

- [図1]図1は、本発明の第1の実施の形態のトランスポート層中継装置の構成図である。
。
- [図2]図2は、本発明の第1の実施の形態のトランスポート層中継装置を使ったネットワーク構成図である。
- [図3]図3は、本発明の第1の実施の形態のトランスポート層中継装置の中継動作タイムチャートである。
- [図4]図4は、本発明の第1の実施の形態のレート制御部の動作フローチャートである。
。
- [図5]図5は、本発明の第2の実施の形態のトランスポート層中継装置の構成図である。
。
- [図6]図6は、本発明の第2の実施の形態のトランスポート層中継装置を使ったネットワーク構成図である。
- [図7]図7は、本発明の第2の実施の形態のレート制御部の動作フローチャートである。
。
- [図8]図8は、本発明の第3の実施の形態のトランスポート層中継装置の構成図である

。

[図9]図9は、本発明の第3の実施の形態のトランスポート層中継装置を使ったネットワーク構成図である。

[図10]図10は、本発明の第3の実施の形態のトランスポート層中継装置の中継動作タイムチャートである。

[図11]図11は、本発明の第3の実施の形態のレート制御部の動作フローチャートである。

[図12]図12は、本発明の第4の実施の形態のトランスポート層中継装置の構成図である。

[図13]図13は、本発明の第4の実施の形態のトランスポート層中継装置を使ったネットワーク構成図である。

[図14]図14は、本発明の第4の実施の形態のレート制御部の動作フローチャートである。

[図15]図15は、本発明の第5の実施の形態のトランスポート層中継装置の構成図である。

[図16]図16は、本発明の第5の実施の形態のトランスポート層中継装置を使ったネットワーク構成図である。

[図17]図17は、本発明の第5の実施の形態のトランスポート層中継装置の中継動作タイムチャートである。

[図18]図18は、本発明の第5の実施の形態のレート制御部の動作フローチャートである。

[図19]図19は、本発明の第6の実施の形態のトランスポート層中継装置の構成図である。

[図20]図20は、本発明の第6の実施の形態のトランスポート層中継装置を使ったネットワーク構成図である。

[図21]図21は、本発明の第6の実施の形態のレート制御部の動作フローチャートである。

[図22]図22は、本発明の第7の実施の形態のトランスポート層中継装置の構成図で

ある。

[図23]図23は、本発明の第7の実施の形態のトランSPORT層中継装置を使ったネットワーク構成図である。

[図24]図24は、本発明の第7の実施の形態のトランSPORT層中継装置の中継動作タイムチャートである。

[図25]図25は、本発明の第7の実施の形態のレート制御部の動作フローチャートである。

[図26]図26は、本発明の第8の実施の形態のトランSPORT層中継装置の構成図である。

[図27]図27は、本発明の第8の実施の形態のトランSPORT層中継装置を使ったネットワーク構成図である。

[図28]図28は、本発明の第8の実施の形態のレート制御部の動作フローチャートである。

[図29]図29は、本発明の第9の実施の形態のトランSPORT層中継装置の構成図である。

[図30]図30は、本発明の第10の実施の形態のトランSPORT層中継装置の構成図である。

発明を実施するための最良の形態

[0034] 次に本発明の実施の形態について図面を参照して詳細に説明する。

[0035] [本発明の第1の実施の形態]

[本発明の第1の実施の形態の構成の説明]

次に本発明の第1の実施の形態の構成について図面を参照して詳細に説明する。

[0036] 図1は、本発明の第1の実施の形態の、トランSPORT層中継装置E001の内部ブロック図である。トランSPORT層中継装置E001は、内部に、中継制御部1001と、中継部1002-1～1002-nと、送信レート制御部1003とを含む。トランSPORT層中継装置E001は、コンピュータによって実現可能である。コンピュータによって実現する場合は、ディスク、半導体メモリ、その他の記録媒体に、中継制御部1001、中継部1002-1～1002-n、送信レート制御部1003を実現するためのプログラムを記録しておく。このブ

ログラムはコンピュータによって読み取られ、その動作を制御することで、コンピュータ上に中継制御部1001、中継部1002-1～1002-n、送信レート制御部1003を実現する。以下、それぞれの構成要素について説明する。

[0037] •中継制御部1001:

中継制御部1001は、端末のトランスポート層のコネクション終端部からのコネクション開設要求、あるいは、他のトランスポート層中継装置の装置間コネクション終端部からの中継要求を受け取り、中継部1002-1～1002-nの中で、現在使われていない、中継部を割り当てて、トランスポート層コネクションの中継処理を行わせる。端末からのコネクション開設要求は、コネクション開設の要求元端末のIPアドレスとポート番号、コネクション開設要求先端末のIPアドレスとポート番号、および、コネクションに使うトランスポートプロトコルの識別子、の5つの情報の組(これをコネクション識別子と呼ぶ)を含む。また、中継要求には、コネクション識別子と共に、トランスポート層中継装置の装置間のコネクション(これを中継コネクションと呼ぶ)を開設するのに必要な情報を含む。

[0038] 中継制御部1001は、中継部1002-1～1002-nのそれぞれが現在使用中かどうかを示す表(空管理表)を持っており、コネクション開設要求、あるいは、中継要求を受けて、中継部を割り当てた時に空管理表の該当中継部を"使用中"とし、中継部から中継終了通知を受け取ると、"非使用中"と識別する情報を書くことによって、どの中継部が現在使用中であるかを管理しているものとする。また、送信レート制御部1003へ、中継部に割り当てているコネクションのコネクション識別子、および、中継部の識別番号(1～n)を登録・削除する。

[0039] 更に、中継制御部1001は、開設要求先端末のIPアドレスと、中継要求の送信先にするトランスポート層中継装置との対応関係を示す表(対応表)を有しており、端末からコネクション開設要求を受け取ったとき、対応表から上記コネクション開設要求に含まれている開設要求先端末のIPアドレスに対応するトランスポート層中継装置を示す情報を取得し、取得した情報を装置間コネクション終端部に渡す。

[0040] •中継部1002-1～1002-n:

内部に端末側コネクション終端部、アプリケーション情報解析部、装置間コネクショ

ン終端部を含む。中継制御部1001から、(1)端末のトランスポート層のコネクション終端部からのコネクション開設要求、あるいは、(2)別のトランスポート層中継装置の装置間コネクション終端部からの中継要求を受け、中継のためのトランスポート層コネクションを、端末、および、他のトランスポート層中継装置との間にそれぞれ開設する。以下、(1)、(2)を受け取った場合、それぞれについて動作を説明する。

[0041] 中継制御部1001から、(1)端末のトランスポート層のコネクション終端部からのコネクション開設要求を受け取った場合、端末側コネクション終端部に、その端末とのトランスポート層コネクションを開設させる。このとき、端末側コネクション終端部は、コネクション開設要求の開設要求先端末の情報(IPアドレスとポート番号、および、コネクションに使うトランスポートプロトコルの識別子)を用い、端末には、コネクション開設要求先の端末と直接接続しているのと同じように見える。また、装置間コネクション終端部に対して、中継制御部1001から渡されている情報によって示されるトランスポート層中継装置(装置Bとする)の装置間コネクション終端部へ中継要求を送信することを指示し、装置Bの装置間コネクション終端部との間にトランスポート層コネクションを開設させる。

[0042] 中継制御部1001から、(2)他のトランスポート層中継装置(装置Bとする)の装置間コネクション終端部からの中継要求を受けた場合、装置間コネクション終端部に、装置Bの装置間コネクション終端部との間でトランスポート層コネクションを開設させる。また、中継要求に含まれる、コネクション識別子の情報を使い、端末側コネクション終端部と、開設要求先端末との間にトランスポート層コネクションを開設する。このとき、端末側コネクション終端部は、コネクション開設要求の開設要求元端末の情報(IPアドレスとポート番号、および、コネクションに使うトランスポートプロトコルの識別子)を用い、端末には、コネクション開設要求元の端末と直接接続しているのと同じように見える。

[0043] 端末側コネクション終端部からコネクション切断通知を受け取ると、装置間コネクション終端部に、中継コネクションを切断させ、中継制御部1001へ中継終了通知を送る。装置間コネクション終端部からコネクション切断通知を受け取った場合も同様に、端末側コネクション終端部に、接続先の端末とのコネクションを切断させ、中継制御部1001へ中継終了通知を送る。

[0044] •端末側コネクション終端部:

トランスポート層コネクションを張った端末との通信に用いる送受信バッファを持ち、トランスポート層コネクションの終端処理を行う。このトランスポート層コネクションの終端処理は、使われるトランスポート層プロトコルに従うものとするが、少なくとも、トランスポート層コネクションの開設、維持、終了の機能、および、フロー制御、輻輳制御機能を持っているものとする。中継部の指示に従い、端末とのコネクション開設・切断を行う。また、端末からコネクション切断要求があれば、コネクション切断後、中継部へコネクション切断通知を送る。

[0045] 端末とコネクションが開設されている間は、端末に対して送信バッファ中のデータを、トランスポート層プロトコルの規約に従って送ると共に、端末からのデータをトランスポート層プロトコルの規約に従って受け取り、受信バッファへと書き込む。また、接続しているアプリケーション情報解析部からデータを受け取り、送信バッファへと書き込む。受信バッファからデータを読み出し、接続しているアプリケーション情報解析部へ受け渡す。

[0046] •アプリケーション情報解析部:

端末側コネクション終端部と、装置間コネクション終端部との間で、トランスポート層のデータの転送を行う。すなわち、端末側コネクション終端部の受信バッファからトランスポート層データを読み出し、装置間コネクション終端部の送信バッファへと書き込む。

[0047] また、装置間コネクション終端部の受信バッファからトランスポート層データを読み出し、端末側コネクション終端部の送信バッファへと書き込む。これらの、トランスポート層データ転送時に、アプリケーション情報を取得し、送信レート制御部1003へと通知する。

[0048] アプリケーション情報とは、たとえば、トランスポート層プロトコル上にHTTP(Hyper Text Transfer Protocol)を使ったデータ送受信が行われている場合、HTTPヘッダ中で、どのようなデータ種別が送られているかを示すContext-typeタグの中身や、送られるデータサイズを示すContext-lengthタグの値である。別の例としては、RTSP(Real Time Streaming Protocol)のコネクションで指定されるストリーミングのレートや、ストリ

ーミングのコネクション識別子などである。また別な例としては、FTPの制御コネクション中で指定される、FTPのデータコネクションのコネクション識別子である。

[0049] •装置間コネクション終端部:

端末側コネクション終端部と同様に、トランスポート層コネクションを張った相手との通信に用いる送受信バッファを持ち、トランスポート層コネクションの終端処理を行う。

[0050] また、中継部の指示に従い、他のトランスポート層中継装置とのコネクション開設・切断を行う。また、他のトランスポート層中継装置からコネクション切断要求があれば、コネクション切断後、中継部へコネクション切断通知を送る。

[0051] 端末側コネクション終端部と異なるのは、トランスポート層コネクションの終端処理に関して、トランスポート層コネクションの開設、維持、終了の機能を持つが、フロー制御、および、輻輳制御のための送信レートの計算は、送信レート制御部が行い、送信レート制御部の指示に従ったレートで送信する点が異なる。また、通信相手からのフィードバック情報をを利用して求めた、現在の通信相手との間のコネクション毎輻輳情報(後述)を送信レート制御部1003に通知する点が異なる。

[0052] また、このトランスポート層コネクションには信頼性あるトランスポート層プロトコルが使われるものとし、再送制御のために、受信側から受信出来たデータのシーケンス番号(ACK番号)を含む、ACKパケットを受け取るものとする。ACK番号は、これまでにロスなく連続して受け取っている最後のシーケンス番号+1の値(RR:Receive Ready)だけでも良いし、SACK(Selective ACK)のように、RRとそれ以降で受け取れているシーケンス番号のブロックを列挙したものであってもよい。さらに、ACKパケットには、受信側の受信バッファの残量などの、フロー制御に必要な情報もふくまれているとする。

[0053] 上記のような、装置間コネクション終端部は、たとえば、TCP Reno、TCP NewReno、あるいは、TCP SACKなどのスライディングウインドウ型の輻輳制御・フロー制御を行うトランスポート層プロトコルのコネクション終端部を一部変更することで実現出来る。具体的には、これらのトランスポートプロトコルでは、送信相手の受信バッファの残量と、輻輳ウインドウの値の小さい方を、最終的な送信ウインドウの値として求めているが、この輻輳ウインドウ(cwndと呼ぶ)を変更している部分を次のように変更する。まず、目標輻輳ウインドウ値(TWと呼ぶ)を決め(決め方は後述)、(a)コネクション開始時、コネク

ションアイドル時(パケットを1RTT以上の時間送っていない時)、あるいはTCPのタイムアウト発生時: cwnd=1とする(b)ACK受信時には、cwnd が TW/2以下なら cwnd = cwnd + 1としそうでなければcwnd = cwnd + 1.0/cwnd とするというように変更する。

[0054] また、目標ウインドウ値が変化した場合、cwndをその目標ウインドウ値で上書きしてもよい。目標ウインドウ値の決め方としては、送信レート制御部が指示するレートを、現在のRTT(往復遅延)で割った値としてもよい。あるいは、装置間コネクション終端部が終端しているTCPコネクションが経由するボトルネックリンクの物理帯域B(bit/秒)と、TCPが観測したRTT(Round Trip Time)の最大値RTT_max(秒)を用いて、 $W_{max}=RTT_{max} \cdot B / (8 \cdot MSS)$ (MSSはTCPの最大セグメントサイズ)とし、送信レート制御部が指示するレート(R (bit/秒)とする)から、 $W_{max} \cdot R / B$ の値を求めてウインドウの値に上書きしてもよい。あるいは、装置間コネクション終端部が終端しているTCPコネクションが経由するボトルネックリンクを流れるクロストラヒックのTCPフローの本数をNとして、 $(4 \cdot N - 1) \cdot W_{max} \cdot R / (4 \cdot N \cdot B - R)$ の値をウインドウの値に上書きしてもよい。

[0055] また、TCPの再送制御に関して通常のTCPプロトコルでは、あるパケットを送ってからタイムアウト時間までACKがこないと再送するという動作に加え、高速再転送と呼ばれる、重複ACKがある閾値(一般的には3という値、高速再転送閾値と呼ぶ)発生するとそのACK番号の次のシーケンス番号のパケットを1回再送するという制御を行っている。本実施の形態の装置間コネクション終端部では、この動作に加え、次のような「再々転送」を行っても良い。再々転送は、高速再転送のきっかけとなった重複ACK番号について、高速再転送後にも引き続いて届く重複ACKの数を数え、これがある閾値(再々転送閾値と呼ぶ)を超えるとふたたび再送するという動作である。再々転送は繰り返し行っても良い。再々転送閾値は、高速再転送を行った時に、これまでに送ったシーケンス番号の最大値から、重複ACK番号と高速再転送閾値の和を引いた値としてもよい。

[0056] コネクションが開設されている間は、トランスポート層プロトコルの規約に従って、コネクションを開設した相手方の装置間コネクション終端部とのトランスポート層コネクションを維持し、送信バッファ中のデータを、(トランスポート層プロトコルの規約に従って)相手方の装置間コネクション終端部へ送ると共に、コネクションを開設した装置間コ

ネクション終端部からのデータを(トランSPORT層プロトコルの規約に従って)受け取り、受信バッファへと書き込む。また、接続しているアプリケーション情報解析部からデータを受け取り、送信バッファへと書き込む。受信バッファからデータを読み出し、接続しているアプリケーション情報解析部へ受け渡す。ただし、送信バッファ中のデータを装置間コネクション終端部へ送る際のレートは、送信レート制御部1003から指示されたレートに従うものとする。また、相手方の装置間コネクション終端部からデータを受け取る際に得られるACKパケットから求まるコネクション毎輻輳情報を送信レート制御部1003へ受け渡すものとする。

[0057] コネクション毎輻輳情報について述べる。一般に、トランSPORT層プロトコルでは、コネクションを張った通信相手からのフィードバック情報として、相手の現在の受信能力(受信バッファの空き量など)と受信したデータのシーケンス番号を含むACKパケットを受け取る。ACKパケットから、現在のパケットの往復遅延(RTT:Round Trip Time)の計測、パケットロスと推測出来るイベントの検知が出来る。RTTは例えば次の様に計算出来る。

[0058] パケット送信時に、そのパケットに含まれるデータのシーケンス番号と、送信時刻を記憶しておく、そのシーケンス番号を受信した旨を知らせるACKパケット受信時に、受信時刻と記憶していた送信時刻との差を求めて得る事が出来る。パケットロスについては、TCPのように、ACKのシーケンス番号が、これまでにロスなく連続して受信出来ている最後のデータのシーケンス番号を返すものであるならば、その重複ACK受信時パケットロスの可能性がある事を検知出来る。また、TCPのFast Retransmitのような、重複ACK受信時に、Go-back-Nではない、高速再送を行う再送アルゴリズムの場合、partial ACK受信時にも、パケットロスの可能性がある事を検知出来る。これらの、RTTやパケットロスの可能性のあるイベントを含む情報を、コネクション毎輻輳情報と呼ぶ。

[0059] •送信レート制御部1003:

現在中継中のコネクションの総送信レートをR_totalとし、これを分割して各中継コネクションに割り当てる。総送信レートR_totalの決め方としては、現在中継中のコネクション数、各アプリケーション情報解析部からのアプリケーション情報、各装置間コネク

ション終端部からのコネクション毎輻輳情報に依存して決定しても良く、例えば次のような方法がある。

[0060] (総送信レート決定方法A1)

所望の実効レート R_{target} に対して、あらかじめ決められた倍率値 α ($\alpha > 1$) を用いて、設定値を $R_{total} = \alpha \cdot R_{target}$ とする。ここで、 $\alpha > 1$ としたのは、パケットロスがあつても所望の実行レート R_{target} でパケットを送信できるようにするためである。

[0061] (総送信レート決定方法B1)

所望の実効レート R_{target} が現在中継中のコネクション数に見合ったものになるよう、 R_{total} を現在中継中のコネクション数に依存して決める。1コネクション当たり R_o のレートを割り当てる決めておき、中継制御部1001によって登録されている現在のコネクション数 m を元に、 $R_{total} = \alpha * R_o * m$ と決定する。 α は、あらかじめ決められた倍率値 α ($\alpha > 1$) である。

[0062] (総送信レート決定方法C1)

所望の実効レート R_{target} が得られるように、各装置間コネクション終端部からのコネクション毎輻輳情報を利用して R_{total} を決める。例えば、時間 T 毎に、総送信レートを更新するものとし、時刻 t に現在 m 本のコネクションを中継しているとし、時刻 $t-T$ から時刻 t の間に、各中継コネクションが C_i ($i=1, \dots, m$) 個のパケットを送り、 L_i ($i=1, \dots, m$) 回のパケットロスを検知し、各中継コネクションの平均パケットサイズが D_i ($i=1, \dots, m$) であったとする。時刻 $t-T$ から時刻 t の間に、

[0063] [数1]

$$\sum_{i=1}^m C_i D_i$$

のデータ量を送ろうとして、

[0064] [数2]

$$\sum_{i=1}^m L_i D_i$$

のデータが、パケットロスで送れなかつたので、データのロス率は、

[0065] [数3]

$$P = \left(\sum_{i=1}^m L_i D_i \right) / \left(\sum_{i=1}^m C_i D_i \right)$$

である。同じ輻輳状況が続くと仮定し、目標レート(R_{target})にするために、

$R_{total} = R_{target} / (1-P)$ とする。ただし、極端な輻輳を避けるため、あるデータロス率の閾値 P_0 に対して、 $P > P_0$ の場合には、 $R_{total} = R_{target} / (1-P_0)$ とする。

[0066] (総送信レート決定方法D1)

出来るだけ大きな実効レートが得られるように、 R_{total} を各装置間コネクション終端部からのコネクション毎輻輳情報をを利用して決める。例えば、時間T毎に、総送信レートを更新するものとし、時刻tに現在m本のコネクションを中継しているとし、時刻t-Tから時刻tの間に、各中継コネクションが C_i ($i=1, \dots, m$)個のパケットを送り、 L_i ($i=1, \dots, m$)回のパケットロスを検知し、各中継コネクションの平均パケットサイズが D_i ($i=1, \dots, m$)であったとする。時刻t-Tから時刻tの間に、

[0067] [数4]

$$\sum_{i=1}^m C_i D_i$$

のデータ量を送ろうとして、

[0068] [数5]

$$\sum_{i=1}^m L_i D_i$$

のデータが、パケットロスで送れなかつたので、データのロス率は、

[0069] [数6]

$$P = \left(\sum_{i=1}^m L_i D_i \right) / \left(\sum_{i=1}^m C_i D_i \right)$$

である。過去の送信レートR_toatlとその時のデータロス率Pから、送信レートと実際に受信側に届いたレート(=R_total * (1-P))の関係を求めた表を作る。この表を作るデータは、順次更新してゆくとする。実際に受信側に届いたレートが最大になる送信レートの値をR_totalとする。ただし、過去の送信レートR_totalがまだ十分に蓄積されていなければ、あらかじめ設定されているR_totalを使う。

[0070] (総送信レート決定方法E1)

現在中継しているアプリケーションに見合った所望の実効レートR_targetが得られるよう、R_totalをアプリケーション情報解析部からの情報も利用して決める。例えば、ストリーミングコネクションのみを中継している時、アプリケーション情報解析部からストリーミングのレート情報を得て、それらの合計値をR_targetとし、送信レート決定方法C1に従ってR_totalを求める。

[0071] (総送信レート決定方法F1)

ボトルネックを共有する他のトラヒックに対する優先度付けをした、実効レートが得られるよう、R_totalを、現在のコネクション数m、および、装置間コネクション終端部からのコネクション毎輻輳情報を利用して決める。例えば、コネクション毎輻輳情報から得られる、中継コネクションが通るネットワークのパケットロス率p、RTTを使ってTCPの1本あたりの実効レートを推定し、その α 倍のレートをm倍してもよい。TCPの1本あたりの実効レートを推定方法としては、M. Handley, S. Floyd, J. Padhye, J. Widmer, "TCP Friendly Rate Control (TFRC): Protocol Specification," Internet Society, RFC 3448, January 2003.の3.1節の"throughput equation"のXで与えられているような計算式でもよい。即ち、中継コネクション終端部からのコネクション毎輻輳情報を利用し、パケットロス発生率p (pの計算の際には、RTT以内のパケットロスも1つのロスイベントと考えて計算する)、RTTの指數平均値R、TCPのタイムアウト時間の推定値t_RTO(例えば、Rの2倍とする)、中継コネクションのあるネットワークのTCPのパケットサイズの平均値をs (バイト) (sは経験的に知られている値でも良い)、b=2として、

[0072] [数7]

$$X = \frac{s}{R \cdot \sqrt{2 \cdot b \cdot p / 3 + (t_RTO \cdot (3 \sqrt{3} \cdot b \cdot p / 8 \cdot p \cdot (1 + 32 \cdot p^2)))}} \quad (\text{bit/sec})$$

と求め、 $R_{\text{total}} = m \cdot \alpha \cdot X$ (bit/sec)と決定する。 α を1より大きくすれば、1本あたりの送信レートを他のトラヒックに対して高くする事が出来、中継するトラヒックの他のトラヒックに対する優先度を相対的に上げることが出来る。逆に、 α を1未満にすれば、1本あたりの送信レートを他のトラヒックに対して低くする事が出来、中継するトラヒックの他のトラヒックに対する優先度を相対的に下げる事が出来る。

[0073] このようにして決めた R_{total} を各中継コネクションへのレートへ割り当てる。割り当ては、現在のコネクション数や、アプリケーション情報解析部からの情報に依存して決めても良く、例えば、次のような方法が考えられる。

[0074] (送信レート割り当て方法1)

全コネクションに公平にレートを割り当てる。即ち、コネクションが m 本あれば、 R_{total}/m のレートを各コネクションに割り当てる。

[0075] (送信レート割り当て方法2)

コネクション識別子のポート番号情報に依存して、決められたルールに従って割り当てる。たとえば、ポート番号毎に優先度を付け、その優先度に応じてレートを割り当てる。

[0076] (送信レート割り当て方法3)

アプリケーション情報解析部からの情報に依存して決定する。例えば、HTTPコネクションでデータ転送を行うアプリケーションの複数のコネクションを中継している場合に、アプリケーション情報解析部から得られた、HTTPのContent-typeフィールドの値に依存して優先度を付けた帯域割り当てを行う。別の例としては、HTTPでファイル取得を行うアプリケーションの複数のコネクションを中継している場合に、アプリケーション情報解析部から得られたHTTPのContent-lengthフィールドの情報に依存して、短いサイズのデータを転送しているコネクションにより高い帯域を割り当てる。また別の例として、Ns本のストリーミングタイプのアプリケーションのコネクションと、Nf本のHTTPを使ったファイル転送タイプのアプリケーションを中継している場合を考える。

アプリケーション情報解析部から得られた、各ストリーミングのレートがRSi($i=1, \dots, N_s$)とし、アプリケーション情報解析部から得られた、ファイル転送タイプのコネクションのファイルサイズをFLi($i=1, \dots, N_f$)とする。ストリーミングタイプのコネクションについては、公平なレート(R_{total} を現在のコネクション数 $N_s + N_f$ で割ったもの)より低いレートのものには、そのストリーミングのレートRSi($i=1, \dots, N_s$)を割り当てる。残ったレート

[0077] [数8]

$$(R_{total} - \sum_{i=1}^{N_s} RS_i)$$

を、まだ割り当てをしていないストリーミングコネクションに優先的に割り当てる、さらに残ったレートを、ファイル転送タイプのコネクションのうちファイルサイズが短いもの順に高いレートを割り当てる。

[0078] [動作の説明]

次に本発明の第1の実施の形態の動作について図面を参照して詳細に説明する。

[0079] 図2は、本発明の第1の実施の形態の、トランスポート層中継装置E001を用いた通信網の構成例である。ネットワークN1の中にはs台の端末H1-1～H1-sがあり、ルータR1を介してネットワークN3と接続されている。ネットワークN2の中には、r台の端末H2-1～H2-rが存在し、ルータR2を介してネットワークN3と接続されている。N1内の端末H1-1～H1-sとN2内の端末H2-1～H2-rが通信するには、ネットワークN3を経由するものとする。また、本発明の第1の実施の形態の、トランスポート層中継装置E001が、ネットワークN1、およびN2に配置されており、それぞれをTR1、TR2と呼ぶ事にする。ルータR1は、端末H1-1～H1-s、あるいは、ネットワークN3からのパケットは全てトランスポート層中継装置TR1へ転送し、トランスポート層中継装置TR1からのパケットについては、通常のルータと同様に転送するものとする。同様に、ルータR2は、端末H2-1～H2-s、あるいは、ネットワークN2からのパケットは全てトランスポート層中継装置TR2へ転送し、トランスポート層中継装置TR2からのパケットについては、通常のルータと同様に転送するものとする。

[0080] 以上の構成の中で、トランスポート層中継装置TR1とルータR1、および、トランスポ

一ト層中継装置TR2とルータR2は同じ装置内にあって、上記と同様の転送を行うようになっていても良い。

[0081] 図3は、図2のネットワークN1内の端末H1-kが、ネットワークN2内の端末H2-jとの間で通信を行う場合の、ネットワークN1、および、ネットワークN2に配置された、第1の実施の形態のトランスポート層中継装置E001(図2、図3の中のTR1、TR2)の動作を示している。端末H1-kは、端末H2-jと通信を行おうとして、コネクション開設要求のパケットを端末H2-jへ向かって送信するが、ルータR1がそのパケットをトランスポート層中継装置TR1へと送り、これをトランスポート層中継装置TR1の中継制御部が受け取る(図3のT100)。

[0082] そして、トランスポート層中継装置TR1内では、中継制御部がそのコネクション開設要求を受け取り、現在使用されていない中継部を見つけて、その中継部の番号とコネクション識別子を送信レート制御部へ登録する(T110)。また、その中継部にコネクション開設要求を受け渡す(T130)。中継部は、内部の端末側コネクション終端部へ指示して(T140)、端末H1-kとの間にコネクションを開設させる(T150)。また、内部の装置間コネクション終端部へ指示して(T160)、トランスポート層中継装置TR2の中継制御部に中継要求を送らせ(T170)、装置間コネクション終端部との間に中継コネクションを開設させる(以下で説明するT180～T210の手順に従う)。中継要求を受け取ったトランスポート層中継装置TR2の中継制御部は、現在使用中でない中継部を見つけて、その中継部の番号と中継要求に含まれるコネクション識別子を送信レート制御部へ登録する(T180)。また、その中継部に中継要求を受け渡す(T190)。中継要求を受け取った中継部は、内部の装置間コネクション終端部へ指示して(T200)、中継要求を出したトランスポート層中継装置TR1の装置間コネクション終端部との間に中継コネクションを開設する(T210)。さらに、中継部は、内部の端末側コネクション終端部へ指示して(T220)、端末H2-jとの間にコネクションを開設させる(以下で説明するT230、T240の手順に従う)。即ち、端末側コネクション終端部が端末H2-jへコネクション開設要求を送り(T230)、コネクション開設を行う(T240)。

[0083] 端末H1-kは、HTTPのGETメソッドを使って、H2-jからファイルを取得するとすれば、TR1とのコネクション開設(T150)の後、HTTPのGETリクエストを送信する(T250)。こ

のリクエストは、トランスポート層中継装置TR1の端末側コネクション終端部が受け取り、アプリケーション情報解析部を経由して装置間コネクション終端部へと受け渡され(T260、T270)、中継コネクションを使って、トランスポート層中継装置TR2の端末間コネクション終端部へと受け渡される(T280)。そして、再び、アプリケーション情報解析部を経由して、端末側コネクション終端部へと渡り(T290、T300)、最終的に、端末H2-jへと届く(T310)。

[0084] 端末H2-jは、このリクエストに応答してデータを端末H1-kへ送り返す(T320)。この応答のデータは、再び、トランスポート層中継装置TR2、および、TR1によって中継されて、端末H1-kへと送り届けられる(T330—T380)。この部分の動作については、詳細を後述する。

[0085] 通信が終わると、どちらかの端末、例えば、端末H2-jがコネクション切断要求を送り(T390)、これを、トランスポート層中継装置TR2の端末側コネクション終端部が受け取り、コネクションを切断する(T400)。端末側コネクション終端部は、中継部に対して、コネクション切断通知を送る(T410)。これを受けた中継部は、装置間コネクション終端部へ、中継終了通知を送り(T420)、装置間コネクション終端部は、接続先との間の中継コネクションを切断する(T430)。また、中継部は、中継制御部へ、中継終了通知を送る(T440)。中継終了通知を受け取った中継制御部は、送信レート制御部へ通知して、コネクション識別子と中継部の番号の登録を削除する(T450)。中継コネクションを切断されたトランスポート層中継装置TR1の装置間コネクション終端部は、中継部へコネクション切断通知を送る(T460)。中継部は、端末側コネクション終端部へ通知して(T470)、接続中の端末(端末H1-k)とのコネクションを切断させる(T480、T490)。その後、中継制御部へ中継終了通知を送り(T500)、それを受けた中継制御部は、送信レート制御部へ通知して、コネクション識別子と中継部の番号の登録を削除する(T510)。

[0086] 以上のような動作により、ネットワークN1内の任意の端末と、ネットワークN2内の任意の端末の間の通信を、本発明の第1の実施例のトランスポート層中継装置TR1、TR2を使って、中継させる事が出来る。

[0087] 上記の動作の中で図3のT320—T380に示したような、ネットワークN2内の端末から

ネットワークN1内の端末へ、データを転送している時の、トランスポート層中継装置の送信レート制御部の動作について詳しく述べる。図4に送信レート制御部の動作のフローチャートを示す。送信レート制御部は、タイマ切れ、あるいは、中継制御部からの中継部の登録あるいは削除、あるいは、アプリケーション情報解析部からのアプリケーション情報通知の、いずれかのイベントの発生によって起動される(F100)。イベントが中継制御部からの登録であった場合(F110)、コネクション識別子と中継部番号を登録し(F120)、もし、登録コネクション数が1であれば(F130)、時間T後に切れるタイマを設定する(F140)。イベントが中継制御部からの削除であった場合(F150)、コネクション識別子、中継部番号、アプリケーション情報(もしあれば)を削除し(F160)、もし、登録コネクション数が0であれば(F170)、タイマを解除する(F180)。イベントがアプリケーション情報解析部からのアプリケーション情報通知であった場合(F190)、そのコネクション識別子に関するアプリケーション情報を登録情報に追加する(F200)。イベントがタイマ切れであった場合(F210)、登録済みの中継部番号に対応する中継部の装置間コネクション終端部からコネクション毎輻輳情報を取得する(F220)。また、現在登録済みのコネクションのアプリケーション情報(F200で追加されているもの)を参照する(F222)。そして、現在登録中のコネクション数も用いて、総送信レートR_totalを決定する(F230)。この決定の方法としては、送信レート制御部1003の説明の中で述べた、総送信レート決定方法A1、B1、C1、D1、E1、F1などを採用することができる。次に、現在登録中の各コネクションの送信レートを決定する(F240)。

[0088] この決定の方法としては、送信レート制御部1003の説明の中で述べた、送信レート割り当て方法1、2、3などを採用することができる。ここで決定した各コネクションの送信レートを、対応する、装置間コネクション終端部へ通知する(F250)。最後に、タイマをTにセットする(F260)。

[0089] **[発明の第1の実施の形態の効果]**

発明の第1の実施の形態で得られる、発明の第1の効果は、複数の端末間のあるトランスポート層コネクション集合に対して、所望の実効レートを確保し、そのコネクション集合の中で、その実効レートを再分配することを、トランスポート層中継装置をネットワークN1およびネットワークN2に一つずつ配置すれば実現出来るため、第1の従来

方法がネットワークN3内の多数のルータを、帯域制御可能なルータに置き換えるのと比べ、装置コスト、および、置き換える手間のコストを低くする事が出来ることである。これは、発明の手法が、中継コネクションのトラヒックと他のトラヒック(クロストラヒック)が同じボトルネックを共有している場合、クロストラヒックを運ぶトранSPORT層プロトコル(たとえばTCP)が輻輳制御を行っていれば、中継コネクションの総送信レートに応じて、送信レートを調整する性質を利用し、総送信レートに応じたある実効レートを確保しているためである。また、総送信レートを、中継しているコネクションの中で分割して利用することで、中継しているあるコネクションの送信レートを下げても、そのレートをクロストラヒックに奪われる事なく、中継している別のコネクションの送信レートとして利用する事が可能となり、実効レートの再配分が可能となっている。

- [0090] 発明の第1の実施の形態で得られる、発明の第2の効果は、送信レート制御部において、総送信レート決定方法A1—E1の方法を用いることにより、第2、第3の従来技術では不可能であった、トランSPORT層コネクション集合に対して、所望の実効レートを確保する事を可能にする事にある。
- [0091] 発明の第1の実施の形態で得られる、発明の第3の効果は、送信レート制御部において、総送信レート決定方法F1の方法を用いることにより、第2、第3の従来技術では不可能であった、端末間の複数のコネクションに対して、他のトラヒックに対して優先度をつけた実効レート得る事を可能にする事である。
- [0092] 発明の第1の実施の形態で得られる、発明の第4の効果は、送信レート割り当て方法3よって、複数の端末間のあるトランSPORT層コネクション集合に対し、ある実効レートを、そのコネクション集合の中で、コネクションを流れるアプリケーションの情報に依存して再配分する事が可能となる。
- [0093] [本発明の第2の実施の形態]
 - [構成の説明]次に本発明の第2の実施の形態の構成について図面を参照して詳細に説明する。
- [0094] 図5は、本発明の第2の実施の形態の、トランSPORT層中継装置E002の内部ブロック図である。本発明の第1の実施の形態のトランSPORT層中継装置E001と比べ、ネットワーク状況推定部2004が加わり、送信レート制御部2003と接続されている点の

みが異なる。

[0095] 中継制御部1001、中継部1002-1～1002-nとその内部の端末側コネクション終端部、アプリケーション情報解析部、装置間コネクション終端部の動作は本発明の第1の実施の形態のトランスポート層中継装置E001のものと同様であるので説明を省略する。以下では、ネットワーク状況推定部2004、および、送信レート制御部2003の説明を行う。なお、トランスポート層中継装置E002は、コンピュータによって実現可能である。コンピュータによって実現する場合は、ディスク、半導体メモリ、その他の記録媒体に、中継制御部1001、中継部1002-1～1002-n、送信レート制御部2003、ネットワーク状況推定部2004を実現するためのプログラムを記録しておく。このプログラムはコンピュータによって読み取られ、その動作を制御することで、コンピュータ上に中継制御部1001、中継部1002-1～1002-n、送信レート制御部2003、ネットワーク状況推定部2004を実現する。

[0096] •ネットワーク状況推定部2004:

ネットワーク状況推定部2004は、ネットワーク内に配置された、他の本発明の第2の実施の形態のトランスポート層中継装置E002のネットワーク状況推定部2004と接続し、計測トラヒックを流すことで、他の本発明の第2の実施の形態のトランスポート層中継装置E002との間のネットワークの状況、たとえば、パケットロス率、片道・往復遅延、ボトルネックの物理帯域や、ボトルネックの空き帯域を推定する。パケットロス率の推定は、例えば、送信側がシーケンス番号を付けた計測パケットを送信し、受信側で到着したパケットのシーケンス番号の抜けを見て、計算結果を送信側へフィードバックするようすればよい。片道遅延については、例えば、トランスポート層中継装置の間で時刻同期を取っておき、送信側から時刻を書き込んだ計測パケットを送り、受信側で到着時間を見て、その差の時間を送信側へフィードバックするようすればよい。往復の遅延は、例えば、計測パケットに対して、受信側が応答パケットを返すようにしておき、計測パケットの送信から応答までの時間を計ることで計測出来る。ボトルネックの物理帯域については、例えば、packet-pairと呼ばれる手法(S.Keshav, B.Landfeldt, A.Seneviratne, B.Melander, P. Gunningberg, "Packet-Pair Flow Control", IEEE/ACM Transactions on Networking, Feb 1995)を使って求めることが

出来る。これは、同じサイズLの計測パケットを連續して(時間間隔を開けずに)送信し、受信側で観測される計測パケットの到着間隔D(あるいは受信側が計測パケットに対して短い長さの応答パケットを返すようにしておき、その応答の到着間隔を送信側で計測してもよい)を観測し、計測パケットの送信側から受信側へのネットワークのボトルネックの物理帶域をL/Dによって求める方法である。ボトルネックの空き帶域については、例えばSLoPSと呼ばれる手法(M.Jain, C.Dovrolis, "End-to-End Available Bandwidth: Measurement Methodology, Dynamics, and Relation with TCP Throughput," Proc. of ACM SIGCOMM, Aug. 2002.)を用いて計測する事が出来る。この方法は、100個程度の計測パケットを連續的にレートRで送ってみてその片方向遅延を見る(これは、送受信者の間で時刻同期などを取れば可能である)手法である。連續したパケットの片方向遅延が段々のびる様で有れば、Rは空き帶域よりも大きいと言え、のびなければ、Rは空き帶域よりも小さいと言える。この判定方法を使い、あるRから初めて、二分探査的に調べれば、空き帶域を知る事が出来る。

[0097] •装置間コネクション終端部:

本発明の第1の装置間コネクション終端部と同じであるが、目標ウインドウ値を決定する際に、装置間コネクション終端部が終端しているTCPコネクションが経由するボトルネックリンクの物理帶域Bの値としてを、ネットワーク状況推定部2004が推定した値を用いてもよい。

[0098] •送信レート制御部2003:

現在中継中のコネクションの総送信レートをR_totalとし、これを分割して、各中継コネクションに割り当てる。総送信レートR_totalの決め方としては、現在中継中のコネクション数、各装置間コネクション終端部からのコネクション毎輻輳情報、アプリケーション情報解析部からのアプリケーション情報、ネットワーク状況推定部2004からの情報に依存して決定しても良く、例えば次のような方法がある。

[0099] (総送信レート決定方法A2)

ネットワーク状況推定部2004から得た情報に基づいて決める。例えば、ネットワーク状況推定部2004から得た、ボトルネック物理帶域がRb、空き帶域がRaであったとする。所望の実効レートR_targetが、ボトルネックの空き帶域Raより小さければ、

$R_{total}=R_{target} \cdot \alpha$ とし、 $R_a < R_{target} < R_b$ であれば、 $R_{total} = R_{target} \cdot \beta$ と決める。 α 、 β は、あらかじめ決められた値であり、 $1 < \alpha$ 、 $1 < \beta$ である。

[0100] (総送信レート決定方法B2)

所望の実効レート R_{target} が現在中継中のコネクション数に見合ったものになるよう、 R_{total} を現在中継中のコネクション数に依存して決める。1コネクション当たり R_o のレートが得られる事を望む場合、中継制御部によって登録されている現在のコネクション数 m を元に、 $R_{target} = R_o * m$ と決定する。所望の帯域 R_{target} が、ボトルネックの空き帯域 R_a より小さければ、 $R_{total}=R_{target} \cdot \alpha$ とし、 $R_a < R_{target} < R_b$ であれば、 $R_{total} = R_{target} \cdot \beta$ と決める。 α 、 β は、あらかじめ決められた値であり、 $1 < \alpha$ 、 $1 < \beta$ である。

[0101] (総送信レート決定方法C2)

各装置間コネクション終端部からのコネクション毎輻輳情報とネットワーク状況推定部2004からの情報を利用して決める。例えば、時間 T 毎に、総送信レートを更新するものとし、時刻 t に現在 m 本のコネクションを中継しているとし、時刻 $t-T$ から時刻 t の間に、各中継コネクションが C_i ($i=1, \dots, m$)個のパケットを送り、 L_i ($i=1, \dots, m$)回のパケットロスを検知したとする。

[0102] $C_1 + C_2 + \dots + C_m < C_{threshold}$ ならば、計測に使うパケット数が不十分で、パケットロス率の精度が悪いので、ネットワーク状況推定部2004から得た、パケットロス率を P とする。そうでなければ、時刻 $t-T$ から時刻 t の間に、各中継コネクションが C_i ($i=1, \dots, m$)個のパケットを送り、 L_i ($i=1, \dots, m$)回のパケットロスを検知し、各中継コネクションの平均パケットサイズが D_i ($i=1, \dots, m$)であったとする。時刻 $t-T$ から時刻 t の間に、

[0103] [数9]

$$\sum_{i=1}^m C_i D_i$$

のデータ量を送ろうとして、

[0104] [数10]

$$\sum_{i=1}^m L_i D_i$$

のデータが、パケットロスで送れなかつたので、データのロス率は、

[0105] [数11]

$$P = \left(\sum_{i=1}^m L_i D_i \right) / \left(\sum_{i=1}^m C_i D_i \right)$$

である。同じ輻輳状況が続くと仮定し、実効レートR_targetを確保するために、

R_total=R_target / (1-P)とする。ただし、極端な輻輳を避けるため、あるパケットロス率の閾値P_0に対して、P > P_0の場合には、R_total=R_target / (1-P_0)とする。

[0106] (総送信レート決定方法D2)

出来るだけ大きな実効レートが得られるように、R_totalを各装置間コネクション終端部からのコネクション毎輻輳情報と、ネットワーク状況推定部2004からの情報とを利用して決める。例えば、時間T毎に、総送信レートを更新するものとし、時刻tに現在m本のコネクションを中継しているとし、時刻t-Tから時刻tの間に、各中継コネクションがC_i (i=1, ..., m)個のパケットを送り、L_i (i=1, ..., m)回のパケットロスを検知したとする。

[0107] $C_1 + C_2 + \dots + C_m < C_{threshold}$ ならば、計測に使うパケット数が不十分で、パケットロス率の精度が悪いので、ネットワーク状況推定部2004から得た、パケットロス率をPとする。そうでなければ、時刻t-Tから時刻tの間に、時刻t-Tから時刻tの間に、各中継コネクションがC_i (i=1, ..., m)個のパケットを送り、L_i (i=1, ..., m)回のパケットロスを検知し、各中継コネクションの平均パケットサイズがD_i (i=1, ..., m)であったとする。時刻t-Tから時刻tの間に、

[0108] [数12]

$$\sum_{i=1}^m C_i D_i$$

のデータ量を送ろうとして、

[0109] [数13]

$$\sum_{i=1}^m L_i D_i$$

のデータが、パケットロスで送れなかつたので、データのロス率は、

[0110] [数14]

$$P = \left(\sum_{i=1}^m L_i D_i \right) / \left(\sum_{i=1}^m C_i D_i \right)$$

である。過去の送信レート R_{total} とその時のデータロス率 P から、送信レートと実際に受信側に届いたレート ($= R_{total} * (1 - P)$) の関係を求めた表を作る。この表を作るデータは、順次更新してゆくとする。実際に受信側に届いたレートが最大になる送信レートの値を R_{total} とする。ただし、過去の送信レート R_{total} がまだ十分に蓄積されていなければ、あらかじめ設定されている R_{total} を使う。

[0111] (総送信レート決定方法E2)

現在中継しているアプリケーションに見合つた所望の実効レート R_{target} が得られるように、 R_{total} をアプリケーション情報解析部からの情報も利用して決める。例えば、ストリーミングコネクションのみを中継している時、アプリケーション情報解析部からストリーミングのレート情報を得て、それらの合計値を R_{target} とし、送信レート決定方法 C2 に従つて R_{total} を求める。

(総送信レート決定方法F2)

ボトルネックを共有する他のトラヒックに対する優先度付けをした、実効レートが得られるように、 R_{total} を、現在のコネクション数 m 、および、装置間コネクション終端部からのコネクション毎輻輳情報と、ネットワーク状況推定部2004からの情報を利用して決める。例えば、コネクション毎輻輳情報、あるいは、ネットワーク状況推定部から得られる、中継コネクションが通るネットワークのパケットロス率 p 、RTT を使って TCP の 1 本あたりの実効レートを推定し、その α 倍のレートを m 倍する。TCP の 1 本あたりの実効レートの推定方法としては、M. Handley, S. Floyd, J. Padhye, J. Widmer, "TCP

Friendly Rate Control (TFRC): Protocol Specification,"Internet Society, RFC 3448, January 2003. の3.1節の"throughput equation"のXで与えられているような計算式でもよい。

[0112] 即ち、中継コネクション終端部からのコネクション毎輻輳情報を利用し、パケットロス発生率p、RTTの指數平均値R、TCPのタイムアウト時間の推定値t_RTO(例えば、Rの2倍とする)、中継コネクションのあるネットワークのTCPのパケットサイズの平均値をs(バイト)(sは経験的に知られている値でも良い)、b=2として、

[0113] [数15]

$$X = \frac{s}{R \cdot \sqrt{2 \cdot b \cdot p / 3} + (t_{RTO} \cdot (3 \sqrt{3 \cdot b \cdot p / 8} \cdot p \cdot (1 + 32 \cdot p^2))))} \quad (\text{bit/sec})$$

と求め、 $R_{\text{total}} = m \cdot \alpha \cdot X$ (bit/sec)と決定する。 α を1より大きくすれば、1本あたりの送信レートを他のトラヒックに対して高くする事が出来、中継するトラヒックの他のトラヒックに対する優先度を相対的に上げることが出来る。逆に、 α を1未満にすれば、1本あたりの送信レートを他のトラヒックに対して低くする事が出来、中継するトラヒックの他のトラヒックに対する優先度を相対的に下げることが出来る。

[0114] このようにして決めた R_{total} を各中継コネクションへ割り当てる。このレート割り当ては、現在のコネクション数や、アプリケーション情報解析部からの情報に依存して決めても良く、例えば、次のような方法が考えられる。

[0115] (送信レート割り当て方法1)

全コネクションに公平にレートを割り当てる。即ち、コネクションがm本あれば、 R_{total}/m のレートを各コネクションに割り当てる。

[0116] (送信レート割り当て方法2)

コネクション識別子のポート番号情報に依存して、決められたルールに従って割り当てる。たとえば、ポート番号毎に優先度を付け、その優先度に応じてレートを割り当てる。

[0117] (送信レート割り当て方法3)

アプリケーション情報解析部からの情報に依存して決定する。例えば、HTTPコネクションでデータ転送を行うアプリケーションの複数のコネクションを中継している場合に、

アプリケーション情報解析部から得られた、HTTPのContent-typeフィールドの値に依存して優先度を付けた帯域割り当てを行う。別の例としては、HTTPでファイル取得を行うアプリケーションの複数のコネクションを中継している場合に、アプリケーション情報解析部から得られたHTTPのContent-lengthフィールドの情報に依存して、短いサイズのデータを転送しているコネクションにより高い帯域を割り当てる。また別の例として、Ns本のストリーミングタイプのアプリケーションのコネクションと、Nf本のHTTPを使ったファイル転送タイプのアプリケーションを中継している場合を考える。アプリケーション情報解析部から得られた、各ストリーミングのレートがRS_i(i=1, …, Ns)とし、アプリケーション情報解析部から得られた、ファイル転送タイプのコネクションのファイルサイズをFL_i(i=1, …, Nf)とする。ストリーミングタイプのコネクションについては、公平なレート(R_{total}を現在のコネクション数Ns+Nfで割ったもの)より低いレートのものには、そのストリーミングのレートRS_i(i=1, …, Ns0)を割り当てる。残ったレート

[0118] [数16]

$$(R_{total} - \sum_{i=1}^{Ns0} RS_i)$$

を、まだ割り当てをしていないストリーミングコネクションに優先的に割り当て、さらに残ったレートを、ファイル転送タイプのコネクションのうちファイルサイズが短いもの順に高いレートを割り当てる。

[0119] [動作の説明]

次に、本発明の第2の実施の形態の動作について説明する。

[0120] 図6は、本発明の第2の実施の形態の、トランスポート層中継装置E002を用いたネットワーク構成の例である。本発明の第1の実施の形態の説明に用いたネットワーク構成の例の図2と比べて、トランスポート層中継装置E001が、本発明の第2の実施の形態の、トランスポート層中継装置E002に置き換わっている点以外は同じ構成である。

[0121] ネットワークN1内の任意の端末と、ネットワークN2内の任意の端末の間の通信を、トランスポート層中継装置TR1、TR2を使って、中継を開始、終了する動作は、図3を用

いて行った、本発明の第1の実施の形態の動作の説明と同じであるため省略する。

[0122] ネットワークN2内の端末からネットワークN1内の端末へ、データを転送している時の、トランスポート層中継装置の送信レート制御部の動作について述べる。図7が送信レート制御部の動作のフローチャートである。送信レート制御部は、タイマ切れ、あるいは、中継制御部からの中継部の登録あるいは削除、あるいは、アプリケーション情報解析部からのアプリケーション情報通知の、いずれかのイベントの発生によって起動される(F100)。

[0123] イベントが中継制御部からの登録であった場合(F110)、コネクション識別子と中継部番号を登録し(F120)、もし、登録コネクション数が1であれば(F130)、時間T後に切れるタイマを設定する(F140)。イベントが中継制御部からの削除であった場合(F150)、コネクション識別子、中継部番号、アプリケーション情報(もしあれば)を削除し(F160)、もし、登録コネクション数が0であれば(F170)、タイマを解除する(F180)。イベントがアプリケーション情報解析部からのアプリケーション情報通知であった場合(F190)、そのコネクション識別子に関するアプリケーション情報を登録情報に追加する(F200)。イベントがタイマ切れであった場合(F210)、登録済みの中継部番号に対応する中継部の装置間コネクション終端部からコネクション毎輻輳情報を取得する(F220)。また、現在登録済みのコネクションのアプリケーション情報(F200で追加されているもの)を参照する(F222)。そして、ネットワーク状況推定部から、ネットワーク状況を取得し(F225)、現在登録中のコネクション数も用いて、総送信レートR_totalを決定する(F230)。この決定の方法としては、送信レート制御部2003の説明の中で述べた、総送信レート決定方法A2、B2、C2、D2、E2、F2等を採用することができる。次に、現在登録中の各コネクションの送信レートを決定する(F240)。この決定の方法としては、送信レート制御部2003の説明の中で述べた、送信レート割り当て方法1、2、3等を採用することができる。ここで決定した各コネクションの送信レートを、対応する、装置間コネクション終端部へ通知する(F250)。最後に、タイマをTにセットする(F260)。

[0124] 【発明の第2の実施の形態の効果】

発明の第2の実施の形態で得られる、発明の第1の効果は、複数の端末間のあるトランスポート層コネクション集合に対して、所望の実効レートを確保し、そのコネクショ

ン集合の中で、その帯域を再分配することを、トランスポート層中継装置をネットワークN1およびネットワークN2に一つずつ配置すれば実現出来たため、第1の従来方法がネットワークN3内の多数のルータを、帯域制御可能なルータに置き換えるのと比べ、装置コスト、および、置き換える手間のコストを低くする事が出来ることである。これは、発明の手法が、中継コネクションのトラヒックと他のトラヒック(クロストラヒック)が同じボトルネックを共有している場合、クロストラヒックを運ぶトランスポート層プロトコル(たとえばTCP)が輻輳制御を行っていれば、中継コネクションの総送信レートに応じて、送信レートを調整する性質を利用し、総送信レートに応じたある実効レートを確保しているためである。また、総送信レートを、中継しているコネクションの中で分割して利用することで、中継しているあるコネクションの送信レートを下げても、そのレートをクロストラヒックに奪われる事なく、中継している別のコネクションの送信レートとして利用する事が可能となり、実効レートの再配分が可能となっている。

[0125] 発明の第2の実施の形態で得られる、発明の第2の効果は、送信レート制御部において、総送信レート決定方法A2～E2の方法を用いることにより、第2、第3の従来技術では不可能であった、トランスポート層コネクション集合に対して、所望の実効レートを確保する事を可能にし、かつ、この送信レート決定方法に関して、ネットワーク状況推定部の情報も併用するため、高い精度で所望の実効レートを確保する事が可能になることにある。

[0126] 発明の第2の実施の形態で得られる、発明の第3の効果は、送信レート制御部において、総送信レート決定方法F2の方法を用いることにより、第2、第3の従来技術では不可能であった、端末間の複数のコネクションに対して、他のトラヒックに対して優先度をつけた帯域を割り当てる事を可能にし、かつ、この送信レート決定方法に関して、ネットワーク状況推定部の情報も併用するため、高い精度で優先度をつけた実効レートを確保する事が可能になることにある。

[0127] 発明の第2の実施の形態で得られる、発明の第4の効果は、送信レート割り当て方法3によって、複数の端末間のあるトランスポート層コネクション集合に対し、ある実効レートを、そのコネクション集合の中で、コネクションを流れるアプリケーションの情報に依存して再配分する事が可能となる。

[0128] [本発明の第3の実施の形態]

[本発明の第3の実施の形態の構成の説明]

次に本発明の第3の実施の形態の構成について図面を参照して詳細に説明する。

[0129] 図8は、本発明の第3の実施の形態の、トランスポート層中継装置E003の内部ブロック図である。トランスポート層中継装置E003は、内部に、中継制御部3001と、中継部3002-1～3002-nと、送信レート制御部3003と、装置間コネクション終端部3004と、MUX-DEMUX部3005とを含む。トランスポート層中継装置E003は、コンピュータによって実現可能である。コンピュータによって実現する場合は、ディスク、半導体メモリ、その他の記録媒体に、中継制御部3001、中継部3002-1～3002-n、送信レート制御部3003、装置間コネクション終端部3004、MUX-DEMUX部3005を実現するためのプログラムを記録しておく。このプログラムはコンピュータによって読み取られ、その動作を制御することで、コンピュータ上に中継制御部3001、中継部3002-1～3002-n、送信レート制御部3003、装置間コネクション終端部3004、MUX-DEMUX部3005を実現する。以下、それぞれの構成要素について説明する。

[0130] • 中継制御部3001:

中継制御部3001は、端末のトランスポート層のコネクション終端部からのコネクション開設要求、対向するトランスポート層中継装置E003の中継開始要求、中継部からのコネクション切断通知、対向するトランスポート層中継装置E003の中継終了要求を受け取る。それを受け取った場合の動作を述べる。

[0131] (端末からのコネクション開設要求を受け取った場合)

端末からのコネクション開設要求は、コネクション開設の要求元端末のIPアドレスとポート番号、コネクション開設要求先端末のIPアドレスとポート番号、および、コネクションに使うトランスポートプロトコルの識別子、の5つの情報の組(これをコネクション識別子と呼ぶ)を含む。コネクション開設要求を受け取った中継制御部は、対向するトランスポート層中継装置E003の中継制御部へ、コネクション開設要求の情報を含んだ中継開始要求を送る。この時、相手方の中継制御部と使われていない中継部(3002-1～3002-n)の番号(1～n)を伝え合い、両方で使われていない中継部番号を決定する。次に、中継制御部は、送信レート制御部に、決定した中継部の識別番号

(1～n)、および、コネクションのコネクション識別子を登録する。そして、MUX-DEMUX部に、決定した中継部番号を登録する。

[0132] さらに、中継制御部は、決定した中継部番号の中継部(3002-1～3002-n)に対して、端末からのコネクション開設を行うように指示する。また装置間コネクション終端部に指示して、もしも、対向するトランスポート層中継装置との間に中継のためのトランスポート層コネクション(中継コネクション)が張られていなければ、コネクションを開設するように指示する。

[0133] (対向するトランスポート層中継装置E003からの中継開始要求を受け取った場合)
対向するトランスポート層中継装置E003からの中継開始要求を受け取ると、相手方の中継制御部と使われていない中継部(3002-1～3002-n)の番号(1～n)を伝え合い、両方で使われていない中継部番号を決定する。次に中継制御部は、送信レート制御部に、決定した中継部の識別番号(1～n)、および、コネクションのコネクション識別子を登録する。そして、MUX-DEMUX部に、決定した中継部番号を登録する。さらに、中継制御部は、決定した中継部番号の中継部(3002-1～3002-n)に対して、中継開始要求に含まれているコネクション開設要求の情報に基づいて、コネクション開設要求先端末とのコネクションの開設を行わせる。

[0134] (中継部からのコネクション切断通知を受け取った場合)
対向するトランスポート層中継装置E003の中継制御部へ、コネクション切断通知を出した中継部の番号を含む、中継終了要求を出す。また、MUX-DEMUX部にコネクション切断通知を出した中継部の中継部番号の登録を削除するよう指示する。さらに、送信レート制御部に、該中継部の中継部番号を伝え、登録を削除する。

[0135] (対向するトランスポート層中継装置E003からの中継終了要求を受け取った場合)
中継終了要求に含まれている中継部番号の中継部に対して、端末とのコネクション切断を指示する。また、MUX-DEMUX部にコネクション切断通知を出した中継部の中継部番号の登録を削除するよう指示する。また、送信レート制御部に、該中継部の中継部番号を伝え、登録を削除する。もし、この時使われている中継部がなくなれば、装置間コネクション終端部に指示して、コネクションを切断するように指示する。

[0136] 中継制御部3001は、中継部3002-1～3002-nのそれぞれが現在使用中かどうかを

示す空管理表を持っており、端末からのコネクション開設要求、あるいは、対向するトランスポート層中継装置E003の中継開始要求中継要求を受けて、中継部を割り当てた時に空管理表の該当中継部を“使用中”とし、中継部からのコネクション切断通知、対向するトランスポート層中継装置E003の中継終了要求を受け取った時に、“非使用中”と識別する情報を書くことによって、どの中継部が現在使用中であるかを管理しているものとする。

[0137] • 中継部3002-1～3002-n:

内部に端末側コネクション終端部、アプリケーション情報解析部を含む。中継制御部からの指示に従って、端末側コネクション終端部に対して、端末との間に、トランスポート層コネクションを開設、あるいは、コネクション切断を行わせる。

• 端末側コネクション終端部:

トランスポート層コネクションを張った端末との通信に用いる送受信バッファを持ち、トランスポート層コネクションの終端処理を行う。このトランスポート層コネクションの終端処理は、使われるトランスポート層プロトコルに従うものとするが、少なくとも、トランスポート層コネクションの開設、維持、終了の機能、および、フロー制御、輻輳制御機能を持っているものとする。中継部の指示に従い、端末とのコネクション開設・切断を行う。また、端末からコネクション切断要求があれば、コネクション切断後、中継部へコネクション切断通知を送る。

[0138] 端末とコネクションが開設されている間は、端末に対して送信バッファ中のデータを、トランスポート層プロトコルの規約に従って端末へ送ると共に、端末からのデータをトランスポート層プロトコルの規約に従って受け取り、受信バッファへと書き込む。また、接続しているアプリケーション情報解析部からデータを受け取り、送信バッファへと書き込む。受信バッファからデータを読み出し、接続しているアプリケーション情報解析部へ受け渡す。

[0139] • アプリケーション情報解析部:

端末側コネクション終端部と、MUX-DEMUX部3005との間で、トランスポート層のデータの転送を行う。すなわち、端末側コネクション終端部の受信バッファからトランスポート層データを読み出し、MUX-DEMUX部3005へ受け渡す。また、MUX-DEMUX

部3005からデータを受け取り、端末側コネクション終端部の送信バッファへと書き込む。これらの、トランスポート層データ転送時に、アプリケーション情報を取得し、送信レート制御部3003へと通知する。アプリケーション情報とは、たとえば、トランスポート層プロトコル上にHTTP(Hyper Text Transfer Protocol)を使ったデータ送受信が行われている場合、HTTPヘッダ中で、どのようなデータ種別が送られているかを示すContext-typeタグの中身や、送られるデータサイズを示すContext-lengthタグの値である。別の例としては、RTSP(Real Time Streaming Protocol)のコネクションで指定されるストリーミングのレートや、ストリーミングのコネクション識別子などである。また別な例としては、FTPの制御コネクション中で指定される、FTPのデータコネクションのコネクション識別子である。

[0140] •MUX-DEMUX部3005:

中継制御部3001から現在使われている中継部の番号の登録・削除の指示に従い、現在使われている中継部の番号を記憶している。現在使われている中継部について、アプリケーション情報解析部からデータを受け取り、ある一定サイズの大きさの断片に分割して(サイズに満たない場合にはパディングして)、中継部番号のついたタグを付けて、装置間コネクション終端部3004の送信バッファへと書き込む。この際、現在使われている中継部のどれから、どれくらいのレートでデータを読み出すかは、送信レート制御部の指示に従う。

[0141] また、装置間コネクション終端部3004の受信バッファから一定サイズ毎にデータの断片を読み出し、その断片についているタグが示す中継部番号の中継部のアプリケーション情報解析部へ、タグを外して断片を受け渡す。

[0142] •装置間コネクション終端部3004:

トランスポート層コネクションを張った相手との通信に用いる送受信バッファを持ち、トランスポート層コネクションの終端処理を行う。また、中継制御部の指示に従い、対向するトランスポート層中継装置E003との間に中継のためのトランスポート層コネクションを張る。

[0143] 端末側コネクション終端部と異なるのは、トランスポート層コネクションの終端処理に関して、トランスポート層コネクションの開設、維持、終了の機能を持つが、フロー制

御、および、輻輳制御のための送信レートの計算は、送信レート制御部が行い、送信レート制御部の指示に従ったレートで送信する点が異なる。また、通信相手からのフィードバック情報をを利用して求めた、現在の通信相手との間のコネクション毎輻輳情報(後述)を送信レート制御部に通知する点が異なる。

[0144] また、このトランスポート層コネクションには信頼性あるトランスポート層プロトコルが使われるものとし、再送制御のために、受信側から受信出来たデータのシーケンス番号(ACK番号)を含む、ACKパケットを受け取るものとする。ACK番号は、これまでにロスなく連続して受け取っている最後のシーケンス番号+1の値(RR:Receive Ready)だけでも良いし、SACK(Selective ACK)のように、RRとそれ以降で受け取られているシーケンス番号のブロックを列挙したものであってもよい。さらに、ACKパケットには、受信側の受信バッファの残量などの、フロー制御に必要な情報もふくまれているとする。

[0145] 上記のような、装置間コネクション終端部は、たとえば、TCP Reno、TCP NewReno、あるいは、TCP SACKなどのスライディングウインドウ型の輻輳制御・フロー制御を行うトランスポート層プロトコルのコネクション終端部を一部変更することで実現出来る。具体的には、これらのトランスポートプロトコルでは、送信相手の受信バッファの残量と、輻輳ウインドウの値の小さい方を、最終的な送信ウインドウの値として求めているが、この輻輳ウインドウ(cwndと呼ぶ)を変更している部分を次のように変更する。まず、目標輻輳ウインドウ値(TWと呼ぶ)を決め(決め方は後述)、(a) コネクション開始時、コネクションアイドル時(パケットを1RTT以上の時間送っていない時)、あるいは、TCPのタイムアウト発生時: $cwnd=1$ とする(b) ACK受信時には、 $cwnd$ が $TW/2$ 以下なら $cwnd = cwnd + 1$ とし、そうでなければ、 $cwnd = cwnd + 1.0/cwnd$ とするというように変更する。

[0146] また、目標ウインドウ値が変化した場合、 $cwnd$ をその目標ウインドウ値で上書きしてもよい。目標ウインドウ値の決め方としては、送信レート制御部が指示するレートを、現在のRTT(往復遅延)で割った値としてもよい。あるいは、装置間コネクション終端部が終端しているTCPコネクションが経由するボトルネックリンクの物理帯域B(bit/秒)と、TCPが観測したRTT(Round Trip Time)の最大値RTT_max(秒)を用いて、 $W_{max}=RTT_{max}\cdot B/(8\cdot MSS)$ (MSSはTCPの最大セグメントサイズ)とし、送信レート制御部が指示するレート(R (bit/秒)とする)から、 $W_{max}\cdot R/B$ の値を求めてウインドウの

値に上書きしてもよい。あるいは装置間コネクション終端部が終端しているTCPコネクションが経由するボトルネックリンクを流れるクロストラヒックのTCPフローの本数をNとして、 $(4 \cdot N - 1) \cdot W_{max} \cdot R / (4 \cdot N \cdot B - R)$ の値をウィンドウの値に上書きしてもよい。

[0147] また、TCPの再送制御に関して、通常のTCPプロトコルでは、あるパケットを送つてからタイムアウト時間までACKがこないと再送するという動作に加え、高速再転送と呼ばれる、重複ACKがある閾値(一般的には3という値、高速再転送閾値と呼ぶ)発生するとそのACK番号の次のシーケンス番号のパケットを1回再送するという制御を行っている。本実施の形態の装置間コネクション終端部では、この動作に加え、次のような「再々転送」を行っても良い。再々転送は、高速再転送のきっかけとなった重複ACK番号について、高速再転送後にも引き続いて届く重複ACKの数を数え、これがある閾値(再々転送閾値と呼ぶ)を超えるとふたたび再送するという動作である、再々転送は繰り返し行つても良い。再々転送閾値は、高速再転送を行つた時に、これまでに送つたシーケンス番号の最大値から、重複ACK番号と高速再転送閾値の和を引いた値としてもよい。

[0148] コネクションが開設されている間は、トランスポート層プロトコルの規約に従つて、コネクションを開設した相手方の装置間コネクション終端部とのトランスポート層コネクションを維持し、送信バッファ中のデータを、(トランスポート層プロトコルの規約に従つて)相手方の装置間コネクション終端部へ送ると共に、コネクションを開設した相手方の装置間コネクション終端部からのデータを(トランスポート層プロトコルの規約に従つて)受け取り、受信バッファへと書き込む。また、接続しているアプリケーション情報解析部からデータを受け取り、送信バッファへと書き込む。受信バッファからデータを読み出し、接続しているアプリケーション情報解析部へ受け渡す。ただし、送信バッファ中のデータを相手方の装置間コネクション終端部へ送る際のレートは、送信レート制御部3003から指示されたレートに従うものとする。また、相手方の装置間コネクション終端部からデータを受け取る際に得られるACKパケットから求まるコネクション毎輻輳情報を送信レート制御部3003へ受け渡すものとする。

[0149] コネクション毎輻輳情報について述べる。一般に、トランスポート層プロトコルでは、コネクションを張った通信相手からのフィードバック情報として、相手の現在の受信能

力(受信バッファの空き量など)と受信したデータのシーケンス番号を含むACKパケットを受け取る。ACKパケットから、現在のパケットの往復遅延(RTT:Round Trip Time)を計測したり、パケットロスと推測出来るイベントを検知することが出来る。RTTは例えば次の様に計算出来る。パケット送信時に、そのパケットに含まれるデータのシーケンス番号と、送信時刻を記憶しておき、そのシーケンス番号を受信した旨を知らせるACKパケット受信時に、受信時刻と記憶していた送信時刻との差を求めて得る事が出来る。パケットロスについては、TCPのように、ACKのシーケンス番号が、これまでにロスなく連続して受信出来ている最後のデータのシーケンス番号を返すものであるならば、その重複ACK受信時パケットロスの可能性がある事を検知出来る。また、TCPのFast Retransmitのような、重複ACK受信時に、Go-back-Nではない、高速再送を行う再送アルゴリズムの場合、partial ACK受信時にも、パケットロスの可能性がある事を検知出来る。これらの、RTTやパケットロスの可能性のあるイベントを含む情報を、コネクション毎輻輳情報と呼ぶ。

[0150] •送信レート制御部3003:

現在中継中のコネクションの総送信レートをR_totalとし、これを中継コネクションの送信レートとする。総送信レートR_totalの決め方としては、現在中継中のコネクション数、装置間コネクション終端部からのコネクション毎輻輳情報に依存して決定しても良く、例えば次のような方法がある。

[0151] (総送信レート決定方法A3)

所望の実効レートR_targetに対して、あらかじめ決められた倍率値 α ($\alpha > 1$)を用いて、設定値を $R_{total} = \alpha \cdot R_{target}$ とする。

[0152] (総送信レート決定方法B3)

所望の実効レートR_targetが現在中継中のコネクション数に見合ったものになるようR_totalを現在中継中のコネクション数に依存して決める。1コネクション当たり R_o のレートを割り当てる決めておき、中継制御部によって登録されている現在のコネクション数mを元に、 $R_{total} = \alpha * R_o * m$ と決定する。 α は、あらかじめ決められた倍率値 α ($\alpha > 1$)である。

[0153] (総送信レート決定方法C3)

所望の実効レート R_{target} が得られるように、装置間コネクション終端部からのコネクション毎輻輳情報を利用して R_{total} を決める。例えば、時間 T 毎に、総送信レートを更新するものとし、時刻 $t-T$ から時刻 t の間に、各中継コネクションが C 個のパケットを送り、 L 回のパケットロスを検知したとすると、パケットロス率は、 $P=L/C$ である。同じ輻輳状況が続くと仮定し、目標レート(R_{target})にするために、 $R_{total}=R_{target}/(1-P)$ とする。ただし、極端な輻輳を避けるため、あるデータロス率の閾値 P_0 に対して、 $P > P_0$ の場合には、 $R_{total}=R_{target}/(1-P_0)$ とする。

[0154] (総送信レート決定方法D3)

出来るだけ大きな実効レートが得られるように、 R_{total} を各装置間コネクション終端部からのコネクション毎輻輳情報を利用して決める。例えば、時間 T 毎に、総送信レートを更新するものとし、時刻 $t-T$ から時刻 t の間に、各中継コネクションが C 個のパケットを送り、 L 回のパケットロスを検知したとすると、パケットロス率は、 $P=L/C$ である。過去の送信レート R_{total} とその時のデータロス率 P から、送信レートと実際に受信側に届いたレート($=R_{total} * (1-P)$)の関係を求めた表を作る。この表を作るデータは、順次更新してゆくとする。

[0155] 実際に受信側に届いたレートが最大になる送信レートの値を R_{total} とする。ただし、過去の送信レート R_{total} がまだ十分に蓄積されていなければ、あらかじめ設定されている R_{total} を使う。

[0156] (総送信レート決定方法E3)

現在中継しているアプリケーションに見合った所望の実効レート R_{target} が得られるように、 R_{total} をアプリケーション情報解析部からの情報も利用して決める。例えば、ストリーミングコネクションのみを中継している時、アプリケーション情報解析部からストリーミングのレート情報を得て、それらの合計値を R_{target} とし、送信レート決定方法C3に従って R_{total} を求める。

[0157] (総送信レート決定方法F3)

ボトルネックを共有する他のトラヒックに対する優先度付けをした、実効レートが得られるように、 R_{total} を、現在のコネクション数 m 、および、装置間コネクション終端部からのコネクション毎輻輳情報を利用して決める。例えば、コネクション毎輻輳情報から

得られる、中継コネクションが通るネットワークのパケットロス率P、RTTを使ってTCPの1本あたりの実効レートを推定し、その α 倍のレートをm倍してもよい。TCPの1本あたりの実効レートの推定方法としては、M. Handley, S. Floyd, J. Padhye, J. Widmer, "TCP Friendly Rate Control(TFRC): Protocol Specification,"Internet Society, RFC 3448, January 2003.の3.1節の"throughput equation"のXで与えられているような計算式でもよい。即ち、中継コネクション終端部からのコネクション毎輻輳情報を利用し、パケットロス発生率p (pの計算の際には、RTT以内のパケットロスも1つのロスイベントと考えて計算する)、RTTの指數平均値R、TCPのタイムアウト時間の推定値t_RTO(例えば、Rの2倍とする)、中継コネクションのあるネットワークのTCPのパケットサイズの平均値をs (バイト) (sは経験的に知られている値でも良い)、b=2として、

[0158] [数17]

$$X = \frac{s}{R \cdot \sqrt{2 \cdot b \cdot p / 3} + (t_{RTO} \cdot (3 \sqrt{3 \cdot b \cdot p / 8} \cdot p \cdot (1 + 32 \cdot p^2)))} \quad (\text{bit/sec})$$

と求め、R_total = m \cdot α \cdot X (bit/sec)と決定する。 α を1より大きくすれば、1本あたりの送信レートを他のトラヒックに対して高くする事が出来、中継するトラヒックの他のトラヒックに対する優先度を相対的に上げることが出来る。逆に、 α を1未満にすれば、1本あたりの送信レートを他のトラヒックに対して低くする事が出来、中継するトラヒックの他のトラヒックに対する優先度を相対的に下げることが出来る。

[0159] このようにして決めたR_totalを各中継コネクションへのレート割り当てでは、現在のコネクション数や、アプリケーション情報解析部からの情報に依存して決めて良く、例えば、次のような方法が考えられる。

[0160] (送信レート割り当て方法1)

全コネクションに公平にレートを割り当てる。即ち、コネクションがm本あれば、R_total/mのレートを各コネクションに割り当てる。

[0161] (送信レート割り当て方法2)

コネクション識別子のポート番号情報に依存して、決められたルールに従って割り当てる。たとえば、ポート番号毎に優先度を付け、その優先度に応じてレートを割り当てる。

[0162] (送信レート割り当て方法3)

アプリケーション情報解析部からの情報に依存して決定する。例えば、HTTPコネクションでデータ転送を行うアプリケーションの複数のコネクションを中継している場合に、アプリケーション情報解析部から得られた、HTTPのContent-typeフィールドの値に依存して優先度を付けた帯域割り当てを行う。別の例としては、HTTPでファイル取得を行うアプリケーションの複数のコネクションを中継している場合に、アプリケーション情報解析部から得られたHTTPのContent-lengthフィールドの情報に依存して、短いサイズのデータを転送しているコネクションにより高い帯域を割り当てる。また別の例として、Ns本のストリーミングタイプのアプリケーションのコネクションと、Nf本のHTTPを使ったファイル転送タイプのアプリケーションを中継している場合を考える。アプリケーション情報解析部から得られた、各ストリーミングのレートがRS_i(i=1, …, N_s)とし、アプリケーション情報解析部から得られた、ファイル転送タイプのコネクションのファイルサイズをFL_i(i=1, …, N_f)とする。ストリーミングタイプのコネクションについては、公平なレート(R_{total}を現在のコネクション数N_s+N_fで割ったもの)より低いレートのものには、そのストリーミングのレートRS_i(i=1, …, N_s)を割り当てる。残ったレート

[0163] [数18]

$$(R_total - \sum_{i=1}^{Ns} RS_i)$$

を、まだ割り当てをしていないストリーミングコネクションに優先的に割り当てる、さらに残ったレートを、ファイル転送タイプのコネクションのうちファイルサイズが短いもの順に高いレートを割り当てる。

[0164] [動作の説明]

次に本発明の第3の実施の形態の動作について図面を参照して詳細に説明する。

[0165] 図9は、本発明の第3の実施の形態の、トランスポート層中継装置E003を用いた通信網の構成例である。ネットワークN1の中にはs台の端末H1-1～H1-sがあり、ルータR1を介してネットワークN3と接続されている。ネットワークN2の中には、r台の端末H2-1～H2-rが存在し、ルータR2を介してネットワークN3と接続されている。N1内の端

末H1-1～H1-sとN2内の端末H2-1～H2-rが通信するには、ネットワークN3を経由するものとする。また、本発明の第3の実施の形態の、トランスポート層中継装置E003が、ネットワークN1、およびN2に配置されており、それぞれをTR1、TR2と呼ぶ事にする。ルータR1は、端末H1-1～H1-s、あるいは、ネットワークN3からのパケットは全てトランスポート層中継装置TR1へ転送し、トランスポート層中継装置TR1からのパケットについては、通常のルータと同様に転送するものとする。同様に、ルータR2は、端末H2-1～H2-s、あるいは、ネットワークN2からのパケットは全てトランスポート層中継装置TR2へ転送し、トランスポート層中継装置TR2からのパケットについては、通常のルータと同様に転送するものとする。

[0166] 以上の構成の中で、トランスポート層中継装置TR1とルータR1、および、トランスポート層中継装置TR2とルータR2は同じ装置内にあって、上記と同様の転送を行うようになっていても良い。

[0167] 図10は、図9のネットワークN1内の端末H1-kが、ネットワークN2内の端末H2-jとの間で通信を行う場合の、ネットワークN1、および、ネットワークN2に配置された、第3の実施の形態のトランスポート層中継装置E003(図9、図10の中のTR1、TR2)の動作を示している。端末H1-kは、端末H2-jと通信を行おうとして、コネクション開設要求のパケットを端末H2-jへ向かって送信するが、ルータR1がそのパケットをトランスポート層中継装置TR1へと送り、これをトランスポート層中継装置TR1の中継制御部が受け取る(図10のT1000)。コネクション開設要求を受け取った、コネクション開設要求を受け取った中継制御部は、対向するトランスポート層中継装置R2の中継制御部へ、コネクション開設要求の情報を含んだ中継開始要求を送る。この時、中継制御部で使われていない中継部(3002-1～3002-n)の番号(1～n)を伝え合い、両方で使われていない中継部番号を決定する(T1010)。

[0168] 次に、送信レート制御部に、決定した中継部の識別番号(1～n)、および、コネクションのコネクション識別子を登録する(T1020)。そして、MUX-DEMUX部に、決定した中継部番号を登録する(T1030)。次に中継制御部は、決定した中継部番号の中継部(3002-1～3002-n)に対して、端末からのコネクション開設を行うように指示する(T1040)。また、T1040でコネクション開設を指示された中継部は、端末側コネクション

終端部に端末とのコネクションを開設させる(T1050、T1060)。

- [0169] また装置間コネクション終端部に指示して、もし、対向するトランスポート層中継装置との間に中継のためのトランスポート層コネクション(中継コネクション)が張られていなければ、コネクションを開設するように指示する(T1070、T1080)。
- [0170] 対向するトランスポート層中継装置R2では、R1からの中継開始要求を受け取ると、中継制御部で使われていない中継部(3002-1～3002-n)の番号(1～n)を伝え合い、両方で使われていない中継部番号を決定する(T1010)。次に、送信レート制御部に、決定した中継部の識別番号(1～n)、および、コネクションのコネクション識別子を登録する(T1090)。そして、MUX-DEMUX部に、決定した中継部番号を登録する(T1100)。次に中継制御部は、決定した中継部番号の中継部(3002-1～3002-n)に対して、中継開始要求に含まれているコネクション開設要求の情報に基づいて、コネクション開設要求先端末とのコネクションの開設を行わせる(T1110、T1120、T1130)。
- [0171] 端末H1-kは、HTTPのGETメソッドを使って、H2-jからファイルを取得するとすれば、TR1とのコネクション開設(T1060)の後、HTTPのGETリクエストを送信する(T1140)。このリクエストは、トランスポート層中継装置TR1の端末側コネクション終端部が受け取り、アプリケーション情報解析部へ受け渡される(T1150)。MUX-DEMUX部は、レート制御部の指示に従ったレートで、アプリケーション情報解析部からデータを受け取り(T1160)、ある一定サイズの大きさの断片に分割して(サイズに満たない場合にはパディングして)、中継部番号のついたタグを付けて、装置間コネクション終端部3004の送信バッファへと書き込む(T1170)。装置間コネクション終端部は、送信バッファ内のデータを、送信レート制御部が指示するレートで送信し、中継コネクションを使って、トランスポート層中継装置TR2の端末間コネクション終端部へと受け渡される(T1180)。トランスポート層中継装置TR2のMUX-DEMUX部は、端末間コネクション終端部の受信バッファから一定サイズ毎にデータの断片を読み出し(T1190)、その断片についているタグが示す中継部番号の中継部のアプリケーション情報解析部へ、タグを外して断片を受け渡す(T1200)。このデータは、アプリケーション情報解析部でを経由して、端末側コネクション終端部へと渡り(T1210)、最終的に、端末H2-jへと届く(T1230)。端末H2-jは、このリクエストに応答してデータを端末H1-kへ送り返す(T1240)。この応答

のデータは、再び、トランスポート層中継装置TR2、および、TR1によって中継されて、端末H1-kへと送り届けられる(T1250～T1320)。この時、TR2のアプリケーション情報解析部は、この応答内のアプリケーション情報、例えば、HTTPのContent-lengthの情報を、送信レート制御部へ受け渡す。この部分の動作については、詳細を後述する。通信が終わると、どちらかの端末、例えば、端末H2-jがコネクション切断要求を送り(T1340)、これを、トランスポート層中継装置TR2の端末側コネクション終端部が受け取り、コネクションを切断する(T1350)。端末側コネクション終端部は、中継部に対して、コネクション切断通知を送る(T1360)。これを受けた中継部は、中継制御部へコネクション切断通知を送る(T1365)。これを受け取った中継制御部は、対向するトランスポート層中継装置TR1の中継制御部へ、コネクション切断通知を出した中継部の番号を含む、中継終了要求を出す(T1370)。また、MUX-DEMUX部に対してコネクション切断通知を出した中継部の中継部番号の登録を削除するよう指示する(T1390)。さらに、送信レート制御部に、該中継部の中継部番号を伝え、登録を削除する(T1400)。中継終了要求を受け取ったトランスポート層中継装置TR1の中継制御部は、中継終了要求に含まれている中継部番号の中継部に対して、端末とのコネクション切断を指示する(T1430)。これを受けた、中継部は、端末側コネクション終端部へ指示して、現在接続中の端末とのコネクションを切断させる(T1440、1450)。また、中継制御部は、MUX-DEMUX部にコネクション切断通知を出した中継部の中継部番号の登録を削除するよう指示する(T1460)。さらに、中継制御部は送信レート制御部に、該中継部の中継部番号を伝え、登録を削除する(T1470)。

- [0172] もし、この時使われている中継部がなくなれば、装置間コネクション終端部に指示して、コネクションを切断するように指示し、コネクションを切断させる(T1480、T1490)。
- [0173] 以上のような動作により、ネットワークN1内の任意の端末と、ネットワークN2内の任意の端末の間の通信を、本発明の第3の実施の形態のトランスポート層中継装置TR1、TR2を使って、中継させる事が出来る。
- [0174] 上記の動作の中で図10のT1240～T1320に示したような、ネットワークN2内の端末からネットワークN1内の端末へ、データを転送している時の、トランスポート層中継装置の送信レート制御部の動作について詳しく述べる。図11に送信レート制御部の動

作のフローチャートを示す。送信レート制御部は、タイマ切れ、あるいは、中継制御部からの中継部の登録あるいは削除、あるいは、アプリケーション情報解析部からのアプリケーション情報通知の、いずれかのイベントの発生によって起動される(F100)。イベントが中継制御部からの登録であった場合(F110)、コネクション識別子と中継部番号を登録し(F120)、もし、登録コネクション数が1であれば(F130)、時間T後に切れるタイマを設定する(F140)。イベントが中継制御部からの削除であった場合(F150)、コネクション識別子、中継部番号、アプリケーション情報(もしあれば)を削除し(F160)、もし、登録コネクション数が0であれば(F170)、タイマを解除する(F180)。イベントがアプリケーション情報解析部からのアプリケーション情報通知であった場合(F190)、そのコネクション識別子に関するアプリケーション情報を登録情報に追加する(F200)。イベントがタイマ切れであった場合(F210)、装置間コネクション終端部からコネクション毎輻輳情報を取得する(F220)。また、現在登録済みのコネクションのアプリケーション情報(F200で追加されているもの)を参照する(F222)。そして、現在登録中のコネクション数も用いて、総送信レートR_totalを決定する(F230)。この決定の方法としては、送信レート制御部3003の説明の中で述べた、総送信レート決定方法A3、B3、C3、D3、E3、F3等を採用することができる。次に、現在登録中の各コネクションの送信レートを決定する(F240)。この決定の方法としては、送信レート制御部3003の説明の中で述べた、送信レート割り当て方法1、2、3等を採用することができる。次に、総送信レートR_totalを装置間コネクション終端部へ通知する(F250)。さらに、F240で決定した各コネクションの送信レートを、MUX-DEMUX部へ通知する(F255)。最後に、タイマをTにセットする(F260)。

[0175] なお、図8に示したトランスポート層中継装置E003は、予め決められた1個の対向するトランスポート層中継装置との間にしか中継コネクションを開設できないが、複数の対向するトランスポート層中継装置との間に中継コネクションを開設できるようにするためには、次のようにすれば良い。

[0176] トランスポート層中継装置E003内に、中継制御部3001を設けると共に、対向する複数のトランスポート層中継装置それぞれに対して、中継部3002-1～3002-n、送信レート制御部3003、装置間コネクション終端部3004およびMUX-DEMUX部3005の組を

設ける。

[0177] 中継制御部3001は、コネクション開設要求を受け取ると、それに含まれているコネクション開設要求先端末のIPアドレスと対応表の内容に基づいて、中継コネクションを開設する相手方のトランスポート層中継装置を決定し、この決定した相手方のトランスポート層中継装置へ中継開始要求を送る。このとき、上記決定した相手方のトランスポート層中継装置に対応する組に含まれる、未使用の中継部の中継部番号を伝え合い、今回使用する中継部の中継部番号を決定する。次に、中継制御部3001は、上記組に属する送信レート制御部3003に、決定した中継部の識別番号および、コネクションのコネクション識別子を登録する。そして、上記組に属するMUX-DEMUX部3005に、決定した中継部番号を登録する。

[0178] さらに、中継制御部3001は、決定した中継部番号の中継部に対して、端末からのコネクション開設を行うように指示する。また、上記組に属する装置間コネクション終端部に指示して、もし、対向するトランスポート層中継装置との間に中継のためのトランスポート層コネクション(中継コネクション)が張られていなければ、コネクションを開設するように指示する。

[0179] また、中継制御部3001は、対向する複数のトランスポート層中継装置の内の或るトランスポート層中継装置から中継開始要求を受け取ると、上記対向するトランスポート層中継装置と対応する組に属する、未使用の中継部の中継部番号を伝え合い、今回使用する中継部番号を決定する。次に中継制御部3001は、上記組の属する送信レート制御部3003に、決定した中継部番号、および、コネクションのコネクション識別子を登録する。そして、上記組に属するMUX-DEMUX部3005に、決定した中継部番号を登録する。さらに、中継制御部3001は、決定した中継部番号の中継部に対して、中継開始要求に含まれているコネクション開設要求の情報に基づいて、コネクション開設要求先端末とのコネクションの開設を行わせる。

[0180] また、中継制御部3001は、ある組に属する中継部からコネクション切断通知を受け取った場合、その組に対応する、対向するトランスポート層中継装置の中継制御部へ、コネクション切断通知を出した中継部の番号を含む、中継終了要求を出す。また、上記組の属するMUX-DEMUX部3005に対して、コネクション切断通知を出した中

継部の中継部番号の登録を削除するよう指示する。さらに、上記組に属する送信レート制御部3003に、該中継部の中継部番号を伝え、登録を削除する。

[0181] また、中継制御部3001は、対向する複数のトランスポート層中継装置の内の或るトランスポート層中継装置から中継終了要求を受け取った場合、上記或るトランスポート層中継装置に対応する組に属する、上記中継終了要求に含まれている中継部番号の中継部に対して、端末とのコネクション切断を指示する。また、上記組に属するMUX-DEMUX部3005に対して、コネクション切断通知を出した中継部の中継部番号の登録を削除するよう指示する。また、上記組の属する送信レート制御部3003に、該中継部の中継部番号を伝え、登録を削除する。もし、この時使われている中継部がなくなれば、装置間コネクション終端部に指示して、コネクションを切断するように指示する。

[0182] **〔発明の第3の実施の形態の効果〕**

発明の第3の実施の形態で得られる、発明の第1の効果は、複数の端末間のあるトランスポート層コネクション集合に対して、所望の実効レートを確保し、そのコネクション集合の中で、その実効レートを再分配することを、トランスポート層中継装置をネットワークN1およびネットワークN2につづつ配置すれば実現出来るため、第1の従来方法がネットワークN3内の多数のルータを、帯域制御可能なルータに置き換えるのと比べ、装置コスト、および、置き換える手間のコストを低くする事が出来ることである。これは、発明の手法が、中継コネクションのトラヒックと他のトラヒック(クロストラヒック)が同じボトルネックを共有している場合、クロストラヒックを運ぶトランスポート層プロトコル(たとえばTCP)が輻輳制御を行っていれば、中継コネクションの総送信レートに応じて、送信レートを調整する性質を利用し、総送信レートに応じたある実効レートを確保しているためである。また、総送信レートを、中継しているコネクションの中で分割して利用することで、中継しているあるコネクションの送信レートを下げても、そのレートをクロストラヒックに奪われる事なく、中継している別のコネクションの送信レートとして利用する事が可能となり、実効レートの再配分が可能となっている。

[0183] 発明の第3の実施の形態で得られる、発明の第2の効果は、送信レート制御部において、総送信レート決定方法A3～E3の方法を用いることにより、第2、第3の従来技

術では不可能であった、トランスポート層コネクション集合に対して、所望の実効レートを確保する事を可能にする事にある。

[0184] 発明の第3の実施の形態で得られる、発明の第3の効果は、送信レート制御部において、総送信レート決定方法F3の方法を用いることにより、第2、第3の従来技術では不可能であった、端末間の複数のコネクションに対して、他のトラヒックに対して優先度をつけた実効レート得る事を可能にする事である。

[0185] 発明の第3の実施の形態で得られる、発明の第4の効果は、送信レート割り当て方法3よって、複数の端末間のあるトランスポート層コネクション集合に対し、ある実効レートを、そのコネクション集合の中で、コネクションを流れるアプリケーションの情報に依存して再配分する事が可能となる。

[0186] [本発明の第4の実施の形態]

[本発明の第4の実施の形態の構成の説明]

次に本発明の第4の実施の形態の構成について図面を参照して詳細に説明する。

[0187] 図12は、本発明の第4の実施の形態の、トランスポート層中継装置E004の内部ブロック図である。本発明の第3の実施の形態のトランスポート層中継装置E003と比べ、ネットワーク状況推定部2004が加わり、送信レート制御部4003と接続されている点のみが異なる。中継制御部3001、中継部3002-1-3002-nとその内部の端末側コネクション終端部、アプリケーション情報解析部、MUX-DEMUX部3005の動作は本発明の第3の実施の形態のトランスポート層中継装置E003のものと同様であるので説明を省略する。また、ネットワーク状況推定部2004の動作も、本発明の第2の実施の形態のトランスポート層中継装置E002のネットワーク状況推定部2004の動作と同じであるため説明を省略する。装置間コネクション終端部3004の動作も、本発明の第3の実施の形態のトランスポート層中継装置E003の装置間コネクション終端部3004と同様であるが、目標ウインドウ値を決定する際に、装置間コネクション終端部が終端しているTCPコネクションが経由するボトルネックリンクの物理帯域Bの値としてを、ネットワーク状況推定部2004が推定した値を用いてもよい。以下では、送信レート制御部4003の説明のみを行う。なお、トランスポート層中継装置E004は、コンピュータによって実現可能である。コンピュータによって実現する場合は、ディスク、半導体メモリ、その他の記録

媒体に、中継制御部3001、中継部3002-1～3002-n、送信レート制御部4003、ネットワーク状況推定部2004、装置間コネクション終端部3004、MUX-DEMUX部3005を実現するためのプログラムを記録しておく。このプログラムはコンピュータによって読み取られ、その動作を制御することで、コンピュータ上に中継制御部3001、中継部3002-1～3002-n、送信レート制御部4003、ネットワーク状況推定部2004、装置間コネクション終端部3004、MUX-DEMUX部3005を実現する。

[0188] ・送信レート制御部4003:

現在中継中のコネクションの総送信レートをR_totalとし、これを、各中継コネクションに割り当てる。総送信レートR_totalの決め方としては、現在中継中のコネクション数、装置間コネクション終端部からのコネクション毎輻輳情報、アプリケーション情報解析部からの情報、ネットワーク状況推定部からの情報に依存して決定しても良く、例えば次のような方法がある。

[0189] (総送信レート決定方法A4)

ネットワーク状況推定部2004から得た情報に基づいて決める。例えば、ネットワーク状況推定部2004から得た、ボトルネック物理帯域Rb、空き帯域がRaであったとする。所望の帯域R_targetが、ボトルネックの空き帯域Raより小さければ、 $R_{total}=R_{target} \cdot \alpha$ とし、 $Ra < R_{target} < Rb$ であれば、 $R_{total} = R_{target} \cdot \beta$ と決める。 α 、 β は、あらかじめ決められた値であり、 $1 < \alpha$ 、 $1 < \beta$ である。

[0190] (総送信レート決定方法B4)

所望の帯域R_targetが現在中継中のコネクション数に見合ったものになるように、 R_{total} を現在中継中のコネクション数に依存して決める。1コネクション当たり R_o のレートが得られる事を望む場合、中継制御部によって登録されている現在のコネクション数mを元に、 $R_{target} = R_o * m$ と決定する。所望の帯域R_targetが、ボトルネックの空き帯域Raより小さければ、 $R_{total}=R_{target} \cdot \alpha$ とし、 $Ra < R_{target} < Rb$ であれば、 $R_{total} = R_{target} \cdot \beta$ と決める。 α 、 β は、あらかじめ決められた値であり、 $1 < \alpha$ 、 $1 < \beta$ である。

[0191] (総送信レート決定方法C4)

中継コネクション終端部からのコネクション毎輻輳情報とネットワーク状況推定部

2004からの情報を利用して決める。例えば、時間T毎に、総送信レートを更新するものとし、時刻 $t-T$ から時刻 t の間に、各中継コネクションが C 個のパケットを送り、 L 回のパケットロスを検知したとする。

- $C < C_{threshold}$ ならば、計測に使うパケット数が不十分で、パケットロス率の精度が悪いので、ネットワーク状況推定部から得た、パケットロス率を P とする。
- そうでなければ、 $P=L/C$ とする。

[0192] 同じ輻輳状況が続くと仮定し、目標レート(R_{target})にするために、 $R_{total}=R_{target}/(1-P)$ とする。ただし、極端な輻輳を避けるため、あるパケットロス率の閾値 P_0 に対して、 $P > P_0$ の場合には、 $R_{total}=R_{target}/(1-P_0)$ とする。

[0193] (総送信レート決定方法D4)

出来るだけ大きな帯域が得られるように、 R_{total} を各装置間コネクション終端部からのコネクション毎輻輳情報とネットワーク状況推定部2004からの情報とを利用して決める。例えば、時間T毎に、総送信レートを更新するものとし、

時刻 $t-T$ から時刻 t の間に、各中継コネクションが C 個のパケットを送り、 L 回のパケットロスを検知したとする。

- $C < C_{threshold}$ ならば、計測に使うパケット数が不十分で、パケットロス率の精度が悪いので、ネットワーク状況推定部から得た、パケットロス率を P とする。
- そうでなければ、 $P=L/C$ とする。

[0194] 過去の送信レート R_{total} とその時のデータロス率 P から、送信レートと実際に受信側に届いたレート($=R_{total} * (1-P)$)の関係を求めた表を作る。この表を作るデータは、順次更新していくとする。実際に受信側に届いたレートが最大になる送信レートの値を R_{total} とする。ただし、過去の送信レート R_{total} がまだ十分に蓄積されていなければ、あらかじめ設定されている R_{total} を使う。

[0195] (総送信レート決定方法E4)

現在中継しているアプリケーションに見合った所望の帯域 R_{target} が得られるように、 R_{total} をアプリケーション情報解析部からの情報も利用して決める。例えば、ストリーミングコネクションのみを中継している時、アプリケーション情報解析部からストリーミングのレート情報を得て、それらの合計値を R_{target} とし、送信レート決定方法C4に

従ってR_totalを求める。

[0196] (総送信レート決定方法F4)

現在のコネクション数m、および、装置間コネクション終端部からのコネクション毎輶情報と、ネットワーク状況推定部2004からの情報を利用して決める。例えば、コネクション毎輶情報、あるいは、ネットワーク状況推定部から得られる、中継コネクションが通るネットワークのパケットロス率p、RTTを使ってTCPの1本あたりの送信レートを推定し、その α 倍のレートをm倍する。TCPの1本あたりの送信レートを推定方法としては、M. Handley, S. Floyd, J. Padhye, J. Widmer, "TCP Friendly Rate Control (TFRC): Protocol Specification," Internet Society, RFC 3448, January 2003. の3.1節の"throughput equation"のXで与えられているような計算式でもよい。即ち、中継コネクション終端部からのコネクション毎輶情報を利用し、パケットロス発生率p、RTTの指指数平均値R、TCPのタイムアウト時間の推定値t_RTO(例えば、Rの2倍とする)、中継コネクションのあるネットワークのTCPのパケットサイズの平均値をs(バイト)(sは経験的に知られている値でも良い)、b=2として、

[0197] [数19]

$$X = \frac{s}{R \cdot \sqrt{2 \cdot b \cdot p / 3} + (t_{RTO} \cdot (3 \sqrt{3 \cdot b \cdot p / 8} \cdot p \cdot (1 + 32 \cdot p^2)))} \quad (\text{bit/sec})$$

と求め、 $R_{total} = m \cdot \alpha \cdot X$ (bit/sec)と決定する。 α を1より大きくすれば、1本あたりの送信レートを他のトラヒックに対して高くする事が出来、中継するトラヒックの他のトラヒックに対する優先度を相対的に上げることが出来る。逆に、 α を1未満にすれば、1本あたりの送信レートを他のトラヒックに対して低くする事が出来、中継するトラヒックの他のトラヒックに対する優先度を相対的に下げる事が出来る。

[0198] このようにして決めたR_totalを各中継コネクションへ割り当てる。このレート割り当ては、現在のコネクション数や、アプリケーション情報解析部からの情報に依存して決めても良く、例えば、次のような方法が考えられる。

[0199] (送信レート割り当て方法1)

全コネクションに公平にレートを割り当てる。即ち、コネクションがm本あれば、 R_{total}/m のレートを各コネクションに割り当てる。

[0200] (送信レート割り当て方法2)

コネクション識別子のポート番号情報に依存して、決められたルールに従って割り当てる。たとえば、ポート番号毎に優先度を付け、その優先度に応じてレートを割り当てる。

[0201] (送信レート割り当て方法3)

アプリケーション情報解析部からの情報に依存して決定する。例えば、HTTPコネクションでデータ転送を行うアプリケーションの複数のコネクションを中継している場合に、アプリケーション情報解析部から得られた、HTTPのContent-typeフィールドの値に依存して優先度を付けた帯域割り当てを行う。別の例としては、HTTPでファイル取得を行うアプリケーションの複数のコネクションを中継している場合に、アプリケーション情報解析部から得られたHTTPのContent-lengthフィールドの情報に依存して、短いサイズのデータを転送しているコネクションにより高い帯域を割り当てる。また別の例として、Ns本のストリーミングタイプのアプリケーションのコネクションと、Nf本のHTTPを使ったファイル転送タイプのアプリケーションを中継している場合を考える。アプリケーション情報解析部から得られた、各ストリーミングのレートがRS_i(i=1, …, Ns)とし、アプリケーション情報解析部から得られた、ファイル転送タイプのコネクションのファイルサイズをFL_i(i=1, …, Nf)とする。ストリーミングタイプのコネクションについては、公平なレート(R_{total}を現在のコネクション数Ns+Nfで割ったもの)より低いレートのものには、そのストリーミングのレートRS_i(i=1, …, Ns)を割り当てる。残ったレート

[0202] [数20]

$$(R_{total} - \sum_{i=1}^{Ns} RS_i)$$

を、まだ割り当てをしていないストリーミングコネクションに優先的に割り当てる、さらに残ったレートを、ファイル転送タイプのコネクションのうちファイルサイズが短いもの順に高いレートを割り当てる。

[0203] [動作の説明]

次に、本発明の第4の実施の形態の動作について説明する。

[0204] 図13は、本発明の第4の実施の形態の、トランスポート層中継装置E004を用いたネットワーク構成の例である。本発明の第3の実施の形態の説明に用いたネットワーク構成の例の図9と比べて、トランスポート層中継装置E003が、本発明の第4の実施の形態の、トランスポート層中継装置E004に置き換わっている点以外は同じ構成である。

[0205] ネットワークN1内の任意の端末と、ネットワークN2内の任意の端末の間の通信を、トランスポート層中継装置TR1、TR2を使って、中継を開始、終了する動作は、図10を用いて行った、本発明の第3の実施の形態の動作の説明と同じであるため省略する。

[0206] ネットワークN2内の端末からネットワークN1内の端末へ、データを転送している時の、トランスポート層中継装置の送信レート制御部の動作について述べる。図14が送信レート制御部の動作のフローチャートである。

[0207] 送信レート制御部は、タイマ切れ、あるいは、中継制御部からの中継部の登録あるいは削除、あるいは、アプリケーション情報解析部からのアプリケーション情報通知の、いずれかのイベントの発生によって起動される(F100)。イベントが中継制御部からの登録であった場合(F110)、コネクション識別子と中継部番号を登録し(F120)、もしも、登録コネクション数が1であれば(F130)、時間T後に切れるタイマを設定する(F140)。イベントが中継制御部からの削除であった場合(F150)、コネクション識別子、中継部番号、アプリケーション情報(もしあれば)を削除し(F160)、もし、登録コネクション数が0であれば(F170)、タイマを解除する(F180)。イベントがアプリケーション情報解析部からのアプリケーション情報通知であった場合(F190)、そのコネクション識別子に関するアプリケーション情報を登録情報に追加する(F200)。イベントがタイマ切れであった場合(F210)、装置間コネクション終端部からコネクション毎輻輳情報を取得する(F220)。また、現在登録済みのコネクションのアプリケーション情報(F200で追加されているもの)を参照する(F222)。そして、ネットワーク状況推定部から、ネットワーク状況を取得し(F225)、現在登録中のコネクション数も用いて、総送信レートR_totalを決定する(F230)。この決定の方法としては、送信レート制御部4003の説明の中で述べた、総送信レート決定方法A4、B4、C4、D4、E4、F4等を採用することができる。次に、現在登録中の各コネクションの送信レートを決定する(F240)。この決定の方法としては、

送信レート制御部4003の説明の中で述べた、送信レート割り当て方法1、2、3等を採用することができる。次に、総送信レートR_totalを装置間コネクション終端部へ通知する(F250)。さらに、F240で決定した各コネクションの送信レートを、MUX-DEMUX部へ通知する(F255)。最後に、タイマをTにセットする(F260)。

[0208] [発明の第4の実施の形態の効果]

発明の第4の実施の形態で得られる、発明の第1の効果は、複数の端末間のあるトランスポート層コネクション集合に対して、所望の実効レートを確保し、そのコネクション集合の中で、その帯域を再分配することを、トランスポート層中継装置をネットワークN1およびネットワークN2に一つずつ配置すれば実現出来るため、第1の従来方法がネットワークN3内の多数のルータを、帯域制御可能なルータに置き換えるのと比べ、装置コスト、および、置き換える手間のコストを低くする事が出来ることである。これは、発明の手法が、中継コネクションのトラヒックと他のトラヒック(クロストラヒック)が同じボトルネックを共有している場合、クロストラヒックを運ぶトランスポート層プロトコル(たとえばTCP)が輻輳制御を行っていれば、中継コネクションの総送信レートに応じて、送信レートを調整する性質を利用し、総送信レートに応じたある実効レートを確保しているためである。また、総送信レートを、中継しているコネクションの中で分割して利用することで、中継しているあるコネクションの送信レートを下げても、そのレートをクロストラヒックに奪われる事なく、中継している別のコネクションの送信レートとして利用する事が可能となり、実効レートの再配分が可能となっている。

[0209] 発明の第4の実施の形態で得られる、発明の第2の効果は、送信レート制御部において、総送信レート決定方法A4～E4の方法を用いることにより、第2、第3の従来技術では不可能であった、トランスポート層コネクション集合に対して、所望の実効レートを確保する事を可能にし、かつ、この送信レート決定方法に関して、ネットワーク状況推定部の情報も併用するため、高い精度で所望の実効レートを確保する事が可能になることがある。

[0210] 発明の第4の実施の形態で得られる、発明の第3の効果は、送信レート制御部において、総送信レート決定方法F4の方法を用いることにより、第2、第3の従来技術では不可能であった、端末間の複数のコネクションに対して、他のトラヒックに対して優先

度をつけた帯域を割り当てるなどを可能にし、かつ、この送信レート決定方法に関して、ネットワーク状況推定部の情報も併用するため、高い精度で優先度をつけた実効レートを確保する事が可能になることがある。

[0211] 発明の第4の実施の形態で得られる、発明の第5の効果は、送信レート割り当て方法3によって、複数の端末間のあるトランスポート層コネクション集合に対し、ある実効レートを、そのコネクション集合の中で、コネクションを流れるアプリケーションの情報に依存して再配分する事が可能となる。

[0212] [本発明の第5の実施の形態]

[本発明の第5の実施の形態の構成の説明]

次に本発明の第5の実施の形態の構成について図面を参照して詳細に説明する。

[0213] 図15は、本発明の第5の実施の形態の、トランスポート層中継装置E005の内部ブロック図である。本発明の第1の実施の形態のトランスポート層中継装置E001と比べ、送信レート制御部5003、中継制御部5001、中継部5002-1～5002-nとその内部の端末側コネクション終端部の動作のみが、変更されている。また、端末のトランスポート層コネクション終端部は、コネクション開設の相手から、コネクション開設時に初期送信レートを与えられると、そのレートで送信を開始し、その後は、通常のレート制御(フロー制御、輻輳制御)に従ったレートで送信を行うものとする。変更部分のみを説明する。なお、トランスポート層中継装置E005は、コンピュータによって実現可能である。コンピュータによって実現する場合は、ディスク、半導体メモリ、その他の記録媒体に、中継制御部5001、中継部5002-1～5002-n、送信レート制御部5003を実現するためのプログラムを記録しておく。

[0214] このプログラムはコンピュータによって読み取られ、その動作を制御することで、コンピュータ上に中継制御部5001、中継部5002-1～5002-n、送信レート制御部5003を実現する。

[0215] •中継制御部5001:

中継制御部5001は、端末のトランスポート層のコネクション終端部からのコネクション開設要求、あるいは、他のトランスポート層中継装置の装置間コネクション終端部からの中継要求を受け取り、中継部5002-1～5002-nの内の現在使われていない中

継部の1つを割り当てる。その後、送信レート制御部5003へ、中継部に割り当てているコネクションのコネクション識別子、および、中継部の識別番号(1～n)を登録する。この際、このコネクションに関して最適な初期送信レート値を送信レート制御部5003から受け取る。この初期送信レート値を、今回割り当てを行った中継部に受け渡すと共に、トランスポート層コネクションの中継処理を行わせる。端末からのコネクション開設要求は、コネクション開設の要求元端末のIPアドレスとポート番号、コネクション開設要求先端末のIPアドレスとポート番号、および、コネクションに使うトランスポートプロトコルの識別子、の5つの情報の組(これをコネクション識別子と呼ぶ)を含む。また、中継要求には、コネクション識別子と共に、トランスポート層中継装置の装置間のコネクション(これを中継コネクションと呼ぶ)を開設するのに必要な情報を含む。

[0216] 中継制御部5001は、中継部5002-1～5002-nのそれぞれが現在使用中かどうかを示す表(空管理表)を持っており、コネクション開設要求、あるいは、中継要求を受けて、中継部を割り当てた時に空管理表の該当中継部を"使用中"とし、中継部から中継終了通知を受け取ると、"非使用中"と識別する情報を書くことによって、どの中継部が現在使用中であるかを管理しているものとする。また、中継部から中継終了通知を受け取ると、送信レート制御部5003へ指示し、中継部に割り当てているコネクションのコネクション識別子、および、中継部の識別番号(1～n)を削除する。

[0217] 更に、中継制御部5001は、開設要求先端末のIPアドレスと、中継要求の送信先にするトランスポート層中継装置との対応関係を示す表(対応表)を有しており、端末からコネクション開設要求を受け取ったとき、対応表から上記コネクション開設要求に含まれている開設要求先端末のIPアドレスに対応するトランスポート層中継装置を示す情報を取得し、取得した情報を装置間コネクション終端部に渡す。

[0218] • 中継部5002-1～5002-n:
内部に端末側コネクション終端部、アプリケーション情報解析部、装置間コネクション終端部を含む。中継制御部から、(1)端末のトランスポート層のコネクション終端部からのコネクション開設要求、あるいは、(2)別のトランスポート層中継装置の装置間コネクション終端部からの中継要求を受け、中継のためのトランスポート層コネクションを、端末、および、他のトランスポート層中継装置との間にそれぞれ開設する。以下、

(1)、(2)を受け取った場合、それぞれについて動作を説明する。

[0219] 中継制御部5001から、(1)端末のトランスポート層のコネクション終端部からのコネクション開設要求を受け取った場合、端末側コネクション終端部に、その端末とのトランスポート層コネクションを開設させる。このとき、中継制御部5001から受け渡された初期送信レート値を端末のトランスポート層のコネクション終端部に伝える。また、このとき、端末側コネクション終端部は、コネクション開設要求の開設要求先端末の情報(IPアドレスとポート番号、および、コネクションに使うトランスポートプロトコルの識別子)を用い、端末には、コネクション開設要求先の端末と直接接続しているのと同じように見える。また、装置間コネクション終端部に対して、中継制御部5001からの情報によって示されるトランスポート層中継装置(例えば、トランスポート層中継装置B)の中継制御部へ中継要求を送信し、トランスポート層中継装置Bの装置間コネクション終端部との間にトランスポート層コネクションを開設することを指示する。

[0220] 中継制御部から、(2)他のトランスポート層中継装置(装置Bとする)の装置間コネクション終端部からの中継要求を受けた場合、装置間コネクション終端部に、装置Bの装置間コネクション終端部との間でトランスポート層コネクションを開設させる。また、中継要求に含まれる、コネクション識別子の情報を使い、端末側コネクション終端部と、開設要求先端末との間にトランスポート層コネクションを開設する。このとき、中継制御部から受け渡された初期送信レート値を端末のトランスポート層のコネクション終端部に伝える。また、このとき、端末側コネクション終端部は、コネクション開設要求の開設要求元端末の情報(IPアドレスとポート番号、および、コネクションに使うトランスポートプロトコルの識別子)を用い、端末には、コネクション開設要求元の端末と直接接続しているのと同じように見える。

[0221] 端末側コネクション終端部からコネクション切断通知を受け取ると、装置間コネクション終端部に、中継コネクションを切断させ、中継制御部へ中継終了通知を送る。装置間コネクション終端部からコネクション切断通知を受け取った場合も同様に、端末側コネクション終端部に、接続先の端末とのコネクションを切断させ、中継制御部へ中継終了通知を送る。

[0222] •端末側コネクション終端部：

トランスポート層コネクションを張った端末との通信に用いる送受信バッファを持ち、トランスポート層コネクションの終端処理を行う。このトランスポート層コネクションの終端処理は、使われるトランスポート層プロトコルに従うものとするが、少なくとも、トランスポート層コネクションの開設、維持、終了の機能、および、フロー制御、輻輳制御機能を持っているものとする。中継部の指示に従い、端末とのコネクション開設・切断を行う。また、端末からコネクション切断要求があれば、コネクション切断後、中継部へコネクション切断通知を送る。ただし、端末とのコネクション開設時には、中継部から受け渡された初期送信レート値を端末のトランスポート層のコネクション終端部に伝える。

[0223] 端末とコネクションが開設されている間は、端末に対して送信バッファ中のデータを、トランスポート層プロトコルの規約に従って端末へ送ると共に、端末からのデータをトランスポート層プロトコルの規約に従って受け取り、受信バッファへと書き込む。また、接続しているアプリケーション情報解析部からデータを受け取り、送信バッファへと書き込む。受信バッファからデータを読み出し、接続しているアプリケーション情報解析部へ受け渡す。

[0224] •送信レート制御部5003:

本発明の実施の形態1の送信レート制御部1003と同様に、現在中継中のコネクションの総送信レートをR_totalとし、これを、分割して各中継コネクションに割り当てる。ただし、中継制御部からのコネクション登録の要求があった時にも、この割り当てを行い、そのコネクションに対する割り当てられた送信レートを、初期送信レートとして中継制御部へ伝える。この動作の詳細は以下で述べる。それ以外の点は、本発明の実施の形態1の送信レート制御部1003と同じである。

[0225] [動作の説明]

次に本発明の第5の実施の形態の動作について図面を参照して説明する。

[0226] 図16は、本発明の第5の実施の形態の、トランスポート層中継装置E005を用いた通信網の構成例である。図2の本発明の第1の実施の形態を用いたネットワーク構成図と比べ、トランスポート層中継装置R1、R2が、図2では、本発明の第1の実施の形態のトランスポート層中継装置E001であったのが、本発明の第5の実施の形態の、トラン

ポート層中継装置E005に置き換わっている点以外は、同じ構成である。

[0227] 図17にネットワークN1内の端末H1-kが、ネットワークN2内の端末H2-jとの間で通信を行う場合の、ネットワークN1、および、ネットワークN2に配置された、第5の実施の形態のトランスポート層中継装置E005の動作を示す。図3を用いて説明した、本発明の第1の実施の形態の動作とほとんど同じなので違いだけを説明する。違いは(1) T110において、送信レート制御部へ登録を行った後、初期送信レートを受け取り(T115)、T130においてそれを中継部へ伝える事。また、(2) 中継部がそれを、T140において、端末側コネクション終端部へ伝える事。さらに、(3) 端末側コネクション終端部はその初期送信レートを、T150のコネクション開設時に端末H1-kへ伝える事。また、(4)T180において、送信レート制御部へ登録を行った後、初期送信レートを受け取り(T185)、T190においてそれを中継部へ伝える事。また、(5) 中継部がそれを、T220において、端末側コネクション終端部へ伝える事。さらに、(6) 端末側コネクション終端部はその初期送信レートを、T240のコネクション開設時に端末H1-kへ伝える事。以上の6点以外は全く同じである。この6点以外は、図3を用いて説明した、本発明の第1の実施の形態の動作と全く同じであるため説明を省略する。

[0228] トランスポート層中継装置E005の送信レート制御部5003の動作についても、変更点を述べる。図18に送信レート制御部の動作のフローチャートを示す。送信レート制御部は、タイマ切れ、あるいは、中継制御部からの中継部の登録あるいは削除、あるいは、アプリケーション情報解析部からのアプリケーション情報通知の、いずれかのイベントの発生によって起動される(F100)。イベントが中継制御部からの登録であった場合(F110)、コネクション識別子と中継部番号を登録し(F120)、新登録フラグを1とし(F121)、F220へ進む。イベントが中継制御部からの削除であった場合(F150)、コネクション識別子、中継部番号、アプリケーション情報(もしあれば)を削除し(F160)、もし、登録コネクション数が0であれば(F170)、タイマを解除する(F180)。イベントがアプリケーション情報解析部からのアプリケーション情報通知であった場合(F190)、そのコネクション識別子に関するアプリケーション情報を登録情報に追加する(F200)。イベントがタイマ切れであった場合(F210)、登録済みの中継部番号に対応する中継部の装置間コネクション終端部からコネクション毎輻輳情報を取得する(F220)。また、現在

登録済みのコネクションのアプリケーション情報(F200で追加されているもの)を参照する(F222)。そして、現在登録中のコネクション数も用いて、総送信レートR_totalを決定する(F230)。この決定の方法としては、送信レート制御部1003の説明の中で述べた、総送信レート決定方法A1、B1、C1、D1、E1、F1などを採用できる。次に、現在登録中の各コネクションの送信レートを決定する(F240)。この決定の方法としては、送信レート制御部1003の説明の中で述べた、送信レート割り当て方法1、2、3などを採用できる。ここで決定した各コネクションの送信レートを、対応する、装置間コネクション終端部へ通知する(F250)。新登録フラグが1であれば、新たに登録されたコネクションに関する送信レートR_iを中継制御部へ通知し(F258)、新登録フラグを0とする(F259)。最後に、タイマをTにセットする(F260)。

[0229] [発明の第5の実施の形態の効果]

発明の第5の実施の形態で得られる発明の効果は、発明の第1の実施の形態で得られる効果に加え、端末とのコネクション設定時に初期送信レートの情報を与えることで、端末がネットワークの状況を学習するまでの時間を短縮し、効率のよい輻輳制御を行わせることが可能になる事である。

[0230] [本発明の第6の実施の形態]

[本発明の第6の実施の形態の構成の説明]

次に本発明の第6の実施の形態の構成について図面を参照して詳細に説明する。

[0231] 図19は、本発明の第6の実施の形態の、トランスポート層中継装置E006の内部ブロック図である。本発明の第2の実施の形態のトランスポート層中継装置E002と比べ、送信レート制御部6003、中継制御部5001、中継部5002-1～5002-nとその内部の端末側コネクション終端部が置き換えられている。また、端末のトランスポート層コネクション終端部は、コネクション開設の相手から、コネクション開設時に初期送信レートを与えられると、そのレートで送信を開始し、その後は、通常のレート制御(フロー制御、輻輳制御)に従ったレートで送信を行うものとする。中継制御部5001、中継部5002-1～5002-nの動作は、本発明の第5の実施の形態で説明しているので、送信レート制御部6003の動作のみ説明する。

[0232] なお、トランスポート層中継装置E006は、コンピュータによって実現可能である。コ

ンピュータによって実現する場合は、ディスク、半導体メモリ、その他の記録媒体に、中継制御部5001、中継部5002-1～5002-n、送信レート制御部6003、ネットワーク状況推定部2004を実現するためのプログラムを記録しておく。このプログラムはコンピュータによって読み取られ、その動作を制御することで、コンピュータ上に中継制御部5001、中継部5002-1～5002-n、送信レート制御部6003、ネットワーク状況推定部2004を実現する。

[0233] ・送信レート制御部6003:

本発明の実施の形態2の送信レート制御部2003と同様に、現在中継中のコネクションの総送信レートをR_totalとし、これを、各中継コネクションに割り当てる。ただし、中継制御部からのコネクション登録の要求があった時にも、この割り当てを行い、そのコネクションに対する割り当てられた送信レートを、初期送信レートとして中継制御部へ伝える。この動作の詳細は以下で述べる。それ以外の点は、本発明の実施の形態2の送信レート制御部2003と同じである。

[0234] 〔動作の説明〕

次に本発明の第6の実施の形態の動作について図面を参照して説明する。

[0235] 図20は、本発明の第6の実施の形態の、トランスポート層中継装置E006を用いた通信網の構成例である。図6の本発明の第2の実施の形態を用いたネットワーク構成図と比べ、トランスポート層中継装置R1、R2が、図20では、本発明の第1の実施の形態のトランスポート層中継装置E002であったのが、本発明の第6の実施の形態の、トランスポート層中継装置E006に置き換わっている点以外は、同じ構成である。

[0236] ネットワークN1内の任意の端末と、ネットワークN2内の任意の端末の間の通信を、トランスポート層中継装置TR1、TR2を使って、中継を開始、終了する動作は、図17を用いて行った、本発明の第5の実施の形態の動作の説明と同じであるため省略する。

[0237] トランスポート層中継装置E006の送信レート制御部の6003動作についても、変更点を述べる。図21に送信レート制御部の動作のフローチャートを示す。送信レート制御部は、タイマ切れ、あるいは、中継制御部からの中継部の登録あるいは削除、あるいは、アプリケーション情報解析部からのアプリケーション情報通知の、いずれかのイベントの発生によって起動される(F100)。イベントが中継制御部からの登録であった場

合(F110)、コネクション識別子と中継部番号を登録し(F120)、新登録フラグを1とし(F121)、F220へ進む。

[0238] イベントが中継制御部からの削除であった場合(F150)、コネクション識別子、中継部番号、アプリケーション情報(もしあれば)を削除し(F160)、もし、登録コネクション数が0であれば(F170)、タイマを解除する(F180)。イベントがアプリケーション情報解析部からのアプリケーション情報通知であった場合(F190)、そのコネクション識別子に関するアプリケーション情報を登録情報に追加する(F200)。イベントがタイマ切れであった場合(F210)、登録済みの中継部番号に対応する中継部の装置間コネクション終端部からコネクション毎輻輳情報を取得する(F220)。また、現在登録済みのコネクションのアプリケーション情報(F200で追加されているもの)を参照する(F222)。そして、ネットワーク状況推定部から、ネットワーク状況を取得し(F225)、現在登録中のコネクション数も用いて、総送信レートR_totalを決定する(F230)。この決定の方法としては、送信レート制御部2003の説明の中で述べた、総送信レート決定方法A2、B2、C2、D2、E2、F2などを採用することができる。次に、現在登録中の各コネクションの送信レートを決定する(F240)。この決定の方法としては、送信レート制御部2003の説明の中で述べた、送信レート割り当て方法1、2、3などを採用することができる。ここで決定した各コネクションの送信レートを、対応する、装置間コネクション終端部へ通知する(F250)。新登録フラグが1であれば、新たに登録されたコネクションに関する送信レートR_iを中継制御部へ通知し(F258)、新登録フラグを0とする(F259)。最後に、タイマをTにセットする(F260)。

[0239] [発明の第6の実施の形態の効果]

発明の第6の実施の形態で得られる発明の効果は、発明の第2の実施の形態で得られる効果に加え、端末とのコネクション設定時に初期送信レートの情報を与えることで、端末がネットワークの状況を学習するまでの時間を短縮し、効率のよい輻輳制御を行わせることができる事である。

[0240] [本発明の第7の実施の形態]

[本発明の第7の実施の形態の構成の説明]

次に本発明の第7の実施の形態の構成について図面を参照して詳細に説明する。

[0241] 図22は、本発明の第7の実施の形態の、トランスポート層中継装置E007の内部ブロック図である。本発明の第3の実施の形態のトランスポート層中継装置E003と比べ、送信レート制御部7003、中継制御部7001、中継部7002-1～7002-nとその内部の端末側コネクション終端部の動作のみが、変更されている。また、端末のトランスポート層コネクション終端部は、コネクション開設の相手から、コネクション開設時に初期送信レートを与えられると、そのレートで送信を開始し、その後は、通常のレート制御(フロー制御、輻輳制御)に従ったレートで送信を行うものとする。変更部分のみを説明する。なお、トランスポート層中継装置E007は、コンピュータによって実現可能である。コンピュータによって実現する場合は、ディスク、半導体メモリ、その他の記録媒体に、中継制御部7001、中継部7002-1～3002-n、送信レート制御部7003、装置間コネクション終端部3004、MUX-DEMUX部3005を実現するためのプログラムを記録しておく。このプログラムはコンピュータによって読み取られ、その動作を制御することで、コンピュータ上に中継制御部7001、中継部7002-1～7002-n、送信レート制御部7003、装置間コネクション終端部3004、MUX-DEMUX部3005を実現する。

[0242] ・中継制御部7001:

中継制御部7001は、端末のトランスポート層のコネクション終端部からのコネクション開設要求、対向するトランスポート層中継装置E007からの中継開始要求、中継部からのコネクション切断通知、対向するトランスポート層中継装置E007からの中継終了要求を受け取る。それぞれを受け取った場合の動作を述べる。

[0243] (端末からのコネクション開設要求を受け取った場合)

端末からのコネクション開設要求は、コネクション開設の要求元端末のIPアドレスとポート番号、コネクション開設要求先端末のIPアドレスとポート番号、および、コネクションに使うトランスポートプロトコルの識別子、の5つの情報の組(これをコネクション識別子と呼ぶ)を含む。コネクション開設要求を受け取った中継制御部7001は、対向するトランスポート層中継装置E007の中継制御部へ、コネクション開設要求の情報を含んだ中継開始要求を送る。この時、中継制御部で使われていない中継部(7002-1～7002-n)の番号(1～n)を伝え合い、両方で使われていない中継部番号を1つ決定する。次に、中継制御部は、送信レート制御部7003に、決定した中継部の識別番号(1

～n)、および、コネクションのコネクション識別子を登録する。この際、このコネクションに関して最適な初期送信レート値を送信レート制御部7003から受け取る。そして、MUX-DEMUX部3005に、決定した中継部番号を登録する。さらに、中継制御部7001は、決定した中継部番号の中継部(7002-1～7002-n)に対して、初期送信レート値を受け渡すと共に、端末からのコネクション開設を行うように指示する。また装置間コネクション終端部3004に指示して、もしも、対向するトランスポート層中継装置との間に中継のためのトランスポート層コネクション(中継コネクション)が張られていなければ、コネクションを開設するように指示する。

[0244] (対向するトランスポート層中継装置E007からの中継開始要求を受け取った場合)

対向するトランスポート層中継装置E007からの中継開始要求を受け取ると、中継制御部同士で使われていない中継部(7002-1～7002-n)の番号(1～n)を伝え合い、両方で使われていない中継部番号を1つ決定する。次に中継制御部7001は、送信レート制御部7003に、決定した中継部の識別番号(1～n)、および、コネクションのコネクション識別子を登録する。この際、このコネクションに関して最適な初期送信レート値を送信レート制御部7003から受け取る。そして、MUX-DEMUX部3005に、決定した中継部番号を登録する。さらに、中継制御部7001は、決定した中継部番号の中継部(7002-1～7002-n)に対して、初期送信レート値を受け渡すと共に、中継開始要求に含まれているコネクション開設要求の情報に基づいて、コネクション開設要求先端末とのコネクションの開設を行わせる。

[0245] (中継部からのコネクション切断通知を受け取った場合)

対向するトランスポート層中継装置E007の中継制御部へ、コネクション切断通知を出した中継部の番号を含む、中継終了要求を出す。また、MUX-DEMUX部3005にコネクション切断通知を出した中継部の中継部番号の登録を削除するよう指示する。さらに、送信レート制御部7003に、該中継部の中継部番号を伝え、登録を削除する。

[0246] (対向するトランスポート層中継装置E007からの中継終了要求を受け取った場合)

中継終了要求に含まれている中継部番号の中継部に対して、端末とのコネクション切断を指示する。また、MUX-DEMUX部3005にコネクション切断通知を出した中継部の中継部番号の登録を削除するよう指示する。また、送信レート制御部7003に、該

中継部の中継部番号を伝え、登録を削除する。もし、この時使われている中継部がなくなれば、装置間コネクション終端部に指示して、コネクションを切断するように指示する。

[0247] 中継制御部7001は、中継部7002-1～7002-nのそれぞれが現在使用中かどうかを示す表(空管理表)を持っており、端末からのコネクション開設要求、あるいは、対向するトランスポート層中継装置E007からの中継開始要求中継要求を受けて、中継部を割り当てた時に空管理表の該当中継部を"使用中"とし、中継部からのコネクション切断通知、対向するトランスポート層中継装置E007からの中継終了要求を受け取った時に、"非使用中"と識別する情報を書くことによって、どの中継部が現在使用中であるかを管理しているものとする。

[0248] •中継部7002-1～7002-n:

内部に端末側コネクション終端部、アプリケーション情報解析部を含む。中継制御部7001からの指示に従って、端末側コネクション終端部に対して、端末との間に、トランスポート層コネクションを開設、あるいは、コネクション切断を行わせる。コネクションを開設の際には、中継制御部7001から受け渡された初期送信レート値を端末のトランスポート層のコネクション終端部に伝える。

[0249] •端末側コネクション終端部:

トランスポート層コネクションを張った端末との通信に用いる送受信バッファを持ち、トランスポート層コネクションの終端処理を行う。このトランスポート層コネクションの終端処理は、使われるトランスポート層プロトコルに従うものとするが、少なくとも、トランスポート層コネクションの開設、維持、終了の機能、および、フロー制御、輻輳制御機能を持っているものとする。中継部の指示に従い、端末とのコネクション開設・切断を行う。また、端末からコネクション切断要求があれば、コネクション切断後、中継部へコネクション切断通知を送る。ただし、端末とのコネクション開設時には、中継部から受け渡された初期送信レート値を端末のトランスポート層のコネクション終端部に伝える。

[0250] 端末とコネクションが開設されている間は、端末に対して送信バッファ中のデータを、トランスポート層プロトコルの規約に従って端末へ送ると共に、端末からのデータをト

トランスポート層プロトコルの規約に従って受け取り、受信バッファへと書き込む。また、接続しているアプリケーション情報解析部からデータを受け取り、送信バッファへと書き込む。受信バッファからデータを読み出し、接続しているアプリケーション情報解析部へ受け渡す。

[0251] •送信レート制御部7003:

現在中継中のコネクションの総送信レートをR_totalとし、これを、中継コネクションの送信レートとする。ただし、中継制御部からのコネクション登録の要求があった時にも、この割り当てを行い、そのコネクションに対する割り当てられた送信レートを、初期送信レートとして中継制御部へ伝える。この動作の詳細は以下で述べる。それ以外の点は、本発明の実施の形態3の送信レート制御部3003と同じである。

[0252] [動作の説明]

次に本発明の第7の実施の形態の動作について図面を参照して詳細に説明する。

[0253] 図23は、本発明の第7の実施の形態の、トランスポート層中継装置E007を用いた通信網の構成例である。図9の本発明の第3の実施の形態を用いたネットワーク構成図と比べ、トランスポート層中継装置R1、R2が、図9では、本発明の第3の実施の形態のトランスポート層中継装置E003であったのが、本発明の第7の実施の形態の、トランスポート層中継装置E007に置き換わっている点以外は、同じ構成である。

[0254] 図24にネットワークN1内の端末H1-kが、ネットワークN2内の端末H2-jとの間で通信を行う場合の、ネットワークN1、および、ネットワークN2に配置された、第7の実施の形態のトランスポート層中継装置E007の動作を示す。図10を用いて説明した、本発明の第3の実施の形態の動作とほとんど同じなので違いだけを説明する。

[0255] 違いは(1) T1020において、送信レート制御部へ登録を行った後、初期送信レートを受け取り(T1025)、T1040においてそれを中継部へ伝える事。また、(2) 中継部がそれを、T1050において、端末側コネクション終端部へ伝える事。さらに、(3) 端末側コネクション終端部はその初期送信レートを、T1060のコネクション開設時に端末H1-kへ伝える事。また、(4) T1090において、送信レート制御部へ登録を行った後、初期送信レートを受け取り(T1095)、T1110においてそれを中継部へ伝える事。また、(5) 中継部がそれを、T1120において、端末側コネクション終端部へ伝える事。さらに、(6) 端

末側コネクション終端部はその初期送信レートを、T1130のコネクション開設時に端末H1-kへ伝える事。以上の6点以外は全く同じである。この6点以外は、図10を用いて説明した、本発明の第3の実施の形態の動作と全く同じであるため説明を省略する。

[0256] トランスポート層中継装置の送信レート制御部7003の動作についても、変更点を述べる。図25に送信レート制御部7003の動作のフローチャートを示す。送信レート制御部は、タイマ切れ、あるいは、中継制御部からの中継部の登録あるいは削除、あるいは、アプリケーション情報解析部からのアプリケーション情報通知の、いずれかのイベントの発生によって起動される(F100)。イベントが中継制御部からの登録であった場合(F110)、コネクション識別子と中継部番号を登録し(F120)、コネクション識別子と中継部番号を登録し(F120)、新登録フラグを1とし(F121)、F220へ進む。イベントが中継制御部からの削除であった場合(F150)、コネクション識別子、中継部番号、アプリケーション情報(もしあれば)を削除し(F160)、もし、登録コネクション数が0であれば(F170)、タイマを解除する(F180)。イベントがアプリケーション情報解析部からのアプリケーション情報通知であった場合(F190)、そのコネクション識別子に関するアプリケーション情報を登録情報に追加する(F200)。イベントがタイマ切れであった場合(F210)、装置間コネクション終端部からコネクション毎輻輳情報を取得する(F220)。また、現在登録済みのコネクションのアプリケーション情報(F200で追加されているもの)を参照する(F222)。そして、現在登録中のコネクション数も用いて、総送信レートR_totalを決定する(F230)。この決定の方法としては、送信レート制御部3003の説明の中で述べた、総送信レート決定方法A3、B3、C3、D3、E3、F3等を採用することができる。次に、現在登録中の各コネクションの送信レートを決定する(F240)。この決定の方法としては、送信レート制御部3003の説明の中で述べた、送信レート割り当て方法1、2、3等を採用することができる。次に、総送信レートR_totalを装置間コネクション終端部へ通知する(F250)。さらに、F240で決定した各コネクションの送信レートを、MUX-DEMUX部へ通知する(F255)。新登録フラグが1であれば、新たに登録されたコネクションに関する送信レートR_iを中継制御部へ通知し(F258)、新登録フラグを0とする(F259)。最後に、タイマをTにセットする(F260)。

[0257] なお、図22に示したトランスポート層中継装置E007は、予め決められた1個の対向

するトランスポート層中継装置との間にしか中継コネクションを開設できないが、複数の対向するトランスポート層中継装置との間に中継コネクションを開設できるようするためには、第3の実施の形態で説明したように、トランスポート層中継装置E007内に、対向する複数のトランスポート層中継装置それぞれに対して、中継部7002-1～7002-n、送信レート制御部7003、装置間コネクション終端部3004およびMUX-DEMUX部3005の組を設け、中継制御部7001に中継制御部3001と同様の動作を行わせれば良い。

[0258] **[発明の第7の実施の形態の効果]**

発明の第7の実施の形態で得られる発明の効果は、発明の第3の実施の形態で得られる効果に加え、端末とのコネクション設定時に初期送信レートの情報を与えることで、端末がネットワークの状況を学習するまでの時間を短縮し、効率のよい輻輳制御を行わせることが可能になる事である。

[0259] **[本発明の第8の実施の形態]**

[本発明の第8の実施の形態の構成の説明]

次に本発明の第8の実施の形態の構成について図面を参照して詳細に説明する。

[0260] 図26は、本発明の第8の実施の形態の、トランスポート層中継装置E008の内部ブロック図である。本発明の第4の実施の形態のトランスポート層中継装置E002と比べ、送信レート制御部8003、中継制御部8001、中継部8002-1～8002-nとその内部の端末側コネクション終端部が置き換えられている。また、端末のトランスポート層コネクション終端部は、コネクション開設の相手から、コネクション開設時に初期送信レートを与えられると、そのレートで送信を開始し、その後は、通常のレート制御(フロー制御、輻輳制御)に従ったレートで送信を行うものとする。中継制御部8001、中継部8002-1～8002-nの動作は、本発明の第7の実施の形態で説明した中継制御部7001、中継部7002-1～7002-nと同様であるので、送信レート制御部8003についてのみ説明する。なお、トランスポート層中継装置E008は、コンピュータによって実現可能である。コンピュータによって実現する場合は、ディスク、半導体メモリ、その他の記録媒体に、中継制御部8001、中継部8002-1～8002-n、送信レート制御部8003、装置間コネクション終端部3004、MUX-DEMUX部3005、ネットワーク状況推定部2004を実現するた

めのプログラムを記録しておく。このプログラムはコンピュータによって読み取られ、その動作を制御することで、コンピュータ上に中継制御部8001、中継部8002-1～8002-n、送信レート制御部8003、装置間コネクション終端部3004、MUX-DEMUX部3005を実現する。

[0261] ・送信レート制御部8003:

現在中継中のコネクションの総送信レートをR_totalとし、これを、中継コネクションの送信レートとする。ただし、中継制御部8001からのコネクション登録の要求があった時にも、この割り当てを行い、そのコネクションに対する割り当てられた送信レートを、初期送信レートとして中継制御部8001へ伝える。この動作の詳細は以下で述べる。それ以外の点は、本発明の実施の形態4の送信レート制御部4003と同じである。

[0262] [動作の説明]

次に本発明の第8の実施の形態の動作について図面を参照して説明する。

[0263] 図27は、本発明の第8の実施の形態の、トランスポート層中継装置E008を用いた通信網の構成例である。図13の本発明の第4の実施の形態を用いたネットワーク構成図と比べ、トランスポート層中継装置R1、R2が、図27では、本発明の第4の実施の形態のトランスポート層中継装置E004であったのが、本発明の第8の実施の形態の、トランスポート層中継装置E008に置き換わっている点以外は、同じ構成である。

[0264] ネットワークN1内の任意の端末と、ネットワークN2内の任意の端末の間の通信を、トランスポート層中継装置TR1、TR2を使って、中継を開始、終了する動作は、図24を用いて行った、本発明の第7の実施の形態の動作の説明と同じであるため省略する。

[0265] トランスポート中継装置E008の送信レート制御部8003の動作についても、変更点を述べる。図28に送信レート制御部8003の動作のフローチャートを示す。

[0266] 送信レート制御部は、タイマ切れ、あるいは、中継制御部からの中継部の登録あるいは削除、あるいは、アプリケーション情報解析部からのアプリケーション情報通知の、いずれかのイベントの発生によって起動される(F100)。イベントが中継制御部からの登録であった場合(F110)、コネクション識別子と中継部番号を登録し(F120)、新登録フラグを1とし(F121)、F220へ進む。イベントが中継制御部からの削除であった場合(F150)、コネクション識別子、中継部番号、アプリケーション情報(もしあれば)を削除

し(F160)、もし、登録コネクション数が0であれば(F170)、タイマを解除する(F180)。イベントがアプリケーション情報解析部からのアプリケーション情報通知であった場合(F190)、そのコネクション識別子に関するアプリケーション情報を登録情報に追加する(F200)。イベントがタイマ切れであった場合(F210)、装置間コネクション終端部からコネクション毎輻輳情報を取得する(F220)。また、現在登録済みのコネクションのアプリケーション情報(F200で追加されているもの)を参照する(F222)。そして、ネットワーク状況推定部から、ネットワーク状況を取得し(F225)、現在登録中のコネクション数も用いて、総送信レートR_totalを決定する(F230)。この決定の方法としては、送信レート制御部4003の説明の中で述べた、総送信レート決定方法A4、B4、C4、D4、F4等を採用できる。次に、現在登録中の各コネクションの送信レートを決定する(F240)。この決定の方法としては、送信レート制御部4003の説明の中で述べた、送信レート割り当て方法1、2、3等を採用できる。次に、総送信レートR_totalを装置間コネクション終端部へ通知する(F250)。さらに、F240で決定した各コネクションの送信レートを、MUX-DEMUX部へ通知する(F255)。新登録フラグが1であれば、新たに登録されたコネクションに関する送信レートR_iを中継制御部へ通知し(F258)、新登録フラグを0とする(F259)。最後に、タイマをTにセットする(F260)。

[0267] なお、図26に示したトランスポート層中継装置E008は、予め決められた1個の対向するトランスポート層中継装置との間にしか中継コネクションを開設できないが、複数の対向するトランスポート層中継装置との間に中継コネクションを開設できるようにするためには、トランスポート層中継装置E008内に、対向する複数のトランスポート層中継装置それぞれに対して、中継部8002-1～8002-n、送信レート制御部8003、装置間コネクション終端部3004、MUX-DEMUX部3005及びネットワーク状況推定部2004の組を設け、中継制御部8001に中継制御部3001と同様の動作を行わせれば良い。

[0268] **[発明の第8の実施の形態の効果]**

発明の第8の実施の形態で得られる発明の効果は、発明の第4の実施の形態で得られる効果に加え、端末とのコネクション設定時に初期送信レートの情報を与えることで、端末がネットワークの状況を学習するまでの時間を短縮し、効率のよい輻輳制御を行わせることが可能になる事である。

[0269] [本発明の第9の実施の形態]

〔構成の説明〕

次に本発明の第9の実施の形態の構成について図面を参照して詳細に説明する。

[0270] 図29は、本発明の第9の実施の形態の、トランスポート層中継装置E009の内部ブロック図である。本発明の第1の実施の形態のトランスポート層中継装置E001と比べ、装置間コネクション終端部へ、インライン計測部が加わっている点、及び送信レート制御部1003の代わりに送信レート制御部9003を備えている点が異なる。インライン計測部の追加によって、本発明の第1の実施の形態から変更が有る部分のみ説明する。なお、トランスポート層中継装置E009は、コンピュータによって実現可能である。コンピュータによって実現する場合は、ディスク、半導体メモリ、その他の記録媒体に、中継制御部1001、中継部1002-1～1002-n、送信レート制御部9003を実現するためのプログラムを記録しておく。このプログラムはコンピュータによって読み取られ、その動作を制御することで、コンピュータ上に中継制御部1001、中継部1002-1～1002-n、送信レート制御部9003を実現する。

[0271] ・インライン計測部:

装置間コネクション終端部に指示し、実際に中継する送信パケットを計測パケットとして利用し、ボトルネックの物理帯域、ボトルネックの空き帯域の情報を推定し、これを、装置間コネクション終端部から、送信レート制御部へのコネクション毎輻輳情報に付与するものである。送信パケットを計測パケットとして利用するボトルネックの物理帯域の推定方法は、例えば、送信パケットを複数個連続して送れる場合に、前述のpacket-pairの手法を使う方法が考えられる。また、送信パケットを計測パケットとして利用したボトルネックの空き帯域の推定方法については、例えば、Cao Le Thanh Man、長谷川剛、村田正幸、"アクティブTCPコネクションを用いたインターネットワーク計測、"電子情報通信学会ソサエティ大会、March 2003、に掲載されている方式を用いればよい。

[0272] ・装置間コネクション終端部:

本発明の第1の実施の形態の装置間コネクション終端部と比べ、(1)インライン計測部からの指示に従ったパケット送信を行う点、インライン計測部へ通信相手からの

ACKパケットの到着とそのヘッダ情報を伝える点、(3)送信レート制御部にコネクション毎輻輳情報を通知する際に、インライン計測部が求めた、ボトルネックの物理帯域、ボトルネックの空き帯域の情報を同時に通知する点のみが異なる。

[0273] ・送信レート制御部9003:

現在中継中のコネクションの総送信レートをR_totalとし、これを、各中継コネクションに割り当てる。総送信レートR_totalの決め方としては、現在中継中のコネクション数、各装置間コネクション終端部からのコネクション毎輻輳情報、アプリケーション情報解析部からの情報に依存して決定しても良く、例えば次のような方法がある。

[0274] (総送信レート決定方法A9)

装置間コネクション終端部からのコネクション毎輻輳情報に依存して決定する。たとえば、各装置間コネクション終端部からのコネクション毎輻輳情報の、ボトルネック物理帯域の推定値、空き帯域の推定値について、各装置間コネクション終端部間での平均値をとて、ボトルネック物理帯域の推定値Rb、空き帯域の推定値Raとする。所望の実効レートR_targetが、ボトルネックの空き帯域Raより小さければ、 $R_{total}=R_{target} \cdot \alpha$ とし、 $Ra < R_{target} < Rb$ であれば、 $R_{total} = R_{target} \cdot \beta$ と決める。 α 、 β は、あらかじめ決められた値であり、 $1 < \alpha$ 、 $1 < \beta$ である。

[0275] (総送信レート決定方法B9)

所望の実効レートR_targetが現在中継中のコネクション数に見合ったものになるよう、R_totalを現在中継中のコネクション数に依存して決める。装置間コネクション終端部からのコネクション毎輻輳情報の、ボトルネック物理帯域の推定値、空き帯域の推定値について、各装置間コネクション終端部間で、尤も推定誤差の小さなものをボトルネック物理帯域の推定値Rb、空き帯域の推定値Raとする(コネクション毎輻輳情報に、計測した回数など、推定誤差に関する情報も付与させればよい)。1コネクション当たりR_oのレートが得られる事を望む場合、中継制御部によって登録されている現在のコネクション数mを元に、 $R_{target} = R_o * m$ と決定する。所望の帯域R_targetが、ボトルネックの空き帯域Raより小さければ、 $R_{total}=R_{target} \cdot \alpha$ とし、 $Ra < R_{target} < Rb$ であれば、 $R_{total}=R_{target} \cdot \beta$ と決める。 α 、 β は、あらかじめ決められた値であり、 $1 < \alpha$ 、 $1 < \beta$ である。

[0276] (総送信レート決定方法C9)

各中継コネクション終端部からのコネクション毎輻輳情報を利用して決める。例えば、時間T毎に、総送信レートを更新するものとし、時刻tに現在m本のコネクションを中継しているとし、時刻t-Tから時刻tの間に、各中継コネクションがC_i (i=1, …, m)個のパケットを送り、L_i (i=1, …, m)回のパケットロスを検知したとする。時刻t-Tから時刻tの間に、時刻t-Tから時刻tの間に、各中継コネクションがC_i (i=1, …, m)個のパケットを送り、L_i (i=1, …, m)回のパケットロスを検知し、各中継コネクションの平均パケットサイズがD_i (i=1, …, m)であったとする。時刻t-Tから時刻tの間に、

[0277] [数21]

$$\sum_{i=1}^m C_i D_i$$

のデータ量を送ろうとして、

[0278] [数22]

$$\sum_{i=1}^m L_i D_i$$

のデータが、パケットロスで送れなかつたので、データのロス率は、

[0279] [数23]

$$P = \left(\sum_{i=1}^m L_i D_i \right) / \left(\sum_{i=1}^m C_i D_i \right)$$

である。同じ輻輳状況が続くと仮定し、実効レートR_targetを確保するために、R_{total}=R_{target} / (1-P)とする。ただし、極端な輻輳を避けるため、あるパケットロス率の閾値P₀に対して、P > P₀の場合には、R_{total}=R_{target} / (1-P₀)とする。

[0280] (総送信レート決定方法D9)

出来るだけ大きな実効レートが得られるように、R_{total}を各装置間コネクション終端部からのコネクション毎輻輳情報を利用して決める。例えば、時間T毎に、総送信レ

ートを更新するものとし、時刻tに現在m本のコネクションを中継しているとし、時刻t-Tから時刻tの間に、各中継コネクションが C_i ($i=1, \dots, m$)個のパケットを送り、 L_i ($i=1, \dots, m$)回のパケットロスを検知したとする。時刻t-Tから時刻tの間に、時刻t-Tから時刻tの間に、各中継コネクションが C_i ($i=1, \dots, m$)個のパケットを送り、 L_i ($i=1, \dots, m$)回のパケットロスを検知し、各中継コネクションの平均パケットサイズが D_i ($i=1, \dots, m$)であったとする。時刻t-Tから時刻tの間に、

[0281] [数24]

$$\sum_{i=1}^m C_i D_i$$

のデータ量を送ろうとして、

[0282] [数25]

$$\sum_{i=1}^m L_i D_i$$

のデータが、パケットロスで送れなかつたので、データのロス率は、

[0283] [数26]

$$P = \left(\sum_{i=1}^m L_i D_i \right) / \left(\sum_{i=1}^m C_i D_i \right)$$

である。過去の送信レート R_{total} とその時のデータロス率Pから、送信レートと実際に受信側に届いたレート($=R_{total} * (1-P)$)の関係を求めた表を作る。この表を作るデータは、順次更新してゆくとする。実際に受信側に届いたレートが最大になる送信レートの値を R_{total} とする。ただし、過去の送信レート R_{total} がまだ十分に蓄積されていなければ、あらかじめ設定されている R_{total} を使う。

[0284] (総送信レート決定方法E9)

現在中継しているアプリケーションに見合った所望の実効レート R_{target} が得られる

ように、R_totalをアプリケーション情報解析部からの情報も利用して決める。例えば、ストリーミングコネクションのみを中継している時、アプリケーション情報解析部からストリーミングのレート情報を得て、それらの合計値をR_targetとし、送信レート決定方法C2に従ってR_totalを求める。

[0285] (総送信レート決定方法F9)

ボトルネックを共有する他のトラヒック(トランスポート層中継装置E001によって中継されるパケットに影響を与える、トランスポート層中継装置E001によって中継されないパケット)に対する優先度付けをした、実効レートが得られるように、R_totalを、現在のコネクション数m、および、装置間コネクション終端部からのコネクション毎輻輳情報を利用して決める。例えば、コネクション毎輻輳情報から得られる、中継コネクションが通るネットワークのパケットロス率p、RTTを使ってTCPの1本あたりの実効レートを推定し、その α 倍のレートをm倍する。TCPの1本あたりの実効レートの推定方法としては、M. Handley, S. Floyd, J. Padhye, J. Widmer, "TCP Friendly Rate Control (TFRC): Protocol Specification," Internet Society, RFC 3448, January 2003. の3. 1節の"throughput equation"のXで与えられているような計算式でもよい。即ち、中継コネクション終端部からのコネクション毎輻輳情報を利用し、パケットロス発生率p、RTTの指指数平均値R、TCPのタイムアウト時間の推定値t_RTO(例えば、Rの2倍とする)、中継コネクションのあるネットワークのTCPのパケットサイズの平均値をs(バイト)(sは経験的に知られている値でも良い)、b=2として、

[0286] [数27]

$$X = \frac{s}{R \cdot \sqrt{2 \cdot b \cdot p / 3 + (t_{RTO} \cdot (3 \sqrt{3 \cdot b \cdot p / 8} \cdot p \cdot (1 + 32 \cdot p^2)))}} \quad (\text{bit/sec})$$

と求め、R_total = m \cdot α \cdot X (bit/sec)と決定する。 α を1より大きくすれば、1本あたりの送信レートを他のトラヒックに対して高くする事が出来、中継するトラヒックの他のトラヒックに対する優先度を相対的に上げることが出来る。逆に、 α を1未満にすれば、1本あたりの送信レートを他のトラヒックに対して低くする事が出来、中継するトラヒックの他のトラヒックに対する優先度を相対的に下げる事が出来る。

[0287] このようにして決めたR_totalを各中継コネクションへ割り当てる。このレート割り当

は、現在のコネクション数や、アプリケーション情報解析部からの情報に依存して決めてても良く、例えば、次のような方法が考えられる。

[0288] (送信レート割り当て方法1)

全コネクションに公平にレートを割り当てる。即ち、コネクションがm本あれば、 R_{total}/m のレートを各コネクションに割り当てる。

[0289] (送信レート割り当て方法2)

コネクション識別子のポート番号情報に依存して、決められたルールに従って割り当てる。たとえば、ポート番号毎に優先度を付け、その優先度に応じてレートを割り当てる。

[0290] (送信レート割り当て方法3)

アプリケーション情報解析部からの情報に依存して決定する。例えば、HTTPコネクションでデータ転送を行うアプリケーションの複数のコネクションを中継している場合に、アプリケーション情報解析部から得られた、HTTPのContent-typeフィールドの値に依存して優先度を付けた帯域割り当てを行う。別の例としては、HTTPでファイル取得を行うアプリケーションの複数のコネクションを中継している場合に、アプリケーション情報解析部から得られたHTTPのContent-lengthフィールドの情報に依存して、短いサイズのデータを転送しているコネクションにより高い帯域を割り当てる。また別の例として、Ns本のストリーミングタイプのアプリケーションのコネクションと、Nf本のHTTPを使ったファイル転送タイプのアプリケーションを中継している場合を考える。アプリケーション情報解析部から得られた、各ストリーミングのレートが RS_i ($i=1, \dots, N_s$)とし、アプリケーション情報解析部から得られた、ファイル転送タイプのコネクションのファイルサイズを FL_i ($i=1, \dots, N_f$)とする。ストリーミングタイプのコネクションについては、公平なレート(R_{total} を現在のコネクション数 $N_s + N_f$ で割ったもの)より低いレートのものには、そのストリーミングのレート RS_i ($i=1, \dots, N_s$)を割り当てる。残ったレート

[0291] [数28]

$$(R_{total} - \sum_{i=1}^{N_s} RS_i)$$

を、まだ割り当てをしていないストリーミングコネクションに優先的に割り当て、さらに残ったレートを、ファイル転送タイプのコネクションのうちファイルサイズが短いもの順に高いレートを割り当てる。

[0292] [動作の説明]

次に、本発明の第9の実施の形態の動作は、本発明の第1の実施の形態の動作と同様である。違いは、装置間コネクション終端部が、その内部のインライン計測部からの、ボトルネックの物理帯域、空き帯域の推定値の情報を付与して、送信レート制御部9003へ、コネクション毎輻輳情報を通知する事と、送信レート制御部が、送信レート決定の際に、A9、B9に示したような、インライン計測部からの、ボトルネックの物理帯域、空き帯域の推定値の情報も用いた送信レートの決定を行う点のみである。

[0293] [発明の第9の実施の形態の効果]

発明の第9の実施の形態では、本発明の第1の実施の形態の効果に加え、次のような効果が得られる。送信レート制御部において、総送信レート決定方法A2、B2の方法を用いることにより、ボトルネックの物理帯域や空き帯域の推定値も考慮した、送信レート決定を行うことにより、高い精度での所望の実効レートの確保や、他のトラヒックに対して高い精度で優先度をつけた実効レートを確保する事が可能になることがある。

[0294] [本発明の第10の実施の形態]

[構成の説明]

次に本発明の第10の実施の形態の構成について図面を参照して詳細に説明する。

[0295] 図30は、本発明の第10の実施の形態の、トランSPORT層中継装置E010の内部ブロック図である。本発明の第3の実施の形態のトランSPORT層中継装置E003と比べ、装置間コネクション終端部へ、インライン計測部が加わっている点、送信レート制御部3003の代わりに送信レート制御部10003を備えている点が異なる。インライン計測部の追加によって、本発明の第3の実施の形態から変更が有る部分のみ説明する。なお、トランSPORT層中継装置E010は、コンピュータによって実現可能である。コンピュータによって実現する場合は、ディスク、半導体メモリ、その他の記録媒体に、中

継制御部3001、中継部3002-1～3002-n、送信レート制御部10003、インライン計測部を有する装置間コネクション終端部10004、MUX-DEMUX部3005を実現するためのプログラムを記録しておく。このプログラムはコンピュータによって読み取られ、その動作を制御することで、コンピュータ上に中継制御部3001、中継部3002-1～3002-n、送信レート制御部10003、インライン計測部を有する装置間コネクション終端部10004、MUX-DEMUX部3005を実現する。

[0296] •インライン計測部:

装置間コネクション終端部に指示し、送信パケットを計測パケットとして利用し、ボトルネックの物理帯域、ボトルネックの空き帯域の情報を推定し、これを、装置間コネクション終端部から、送信レート制御部へのコネクション毎輻輳情報に付与するものである。

[0297] 送信パケットを計測パケットとして利用するボトルネックの物理帯域の推定方法は、例えば、送信パケットを複数個連続して送れる場合に、前述のpacket-pairの手法を使う方法が考えられる。また、送信パケットを計測パケットとして利用したボトルネックの空き帯域の推定方法については、例えば、Cao Le Thanh Man、長谷川剛、村田正幸、"アクティブTCPコネクションを用いたインターネットワーク計測、"電子情報通信学会ソサエティ大会、March 2003、に掲載されている方式を用いればよい。

[0298] •装置間コネクション終端部10004:

本発明の第3の実施の形態の装置間コネクション終端部と比べ、(1)インライン計測部からの指示に従ったパケット送信を行う点、インライン計測部へ通信相手からのACKパケットの到着とそのヘッダ情報を伝える点、(3)送信レート制御部にコネクション毎輻輳情報を通知する際に、インライン計測部が求めた、ボトルネックの物理帯域、ボトルネックの空き帯域の情報を同時に通知する点のみが異なる。

[0299] •送信レート制御部10003:

現在中継中のコネクションの総送信レートをR_totalとし、これを、各中継コネクションに割り当てる。総送信レートR_totalの決め方としては、総送信レートR_totalの決め方としては、現在中継中のコネクション数、各装置間コネクション終端部からのコネクション毎輻輳情報、アプリケーション情報解析部からの情報に依存して決定しても良く、

例えば次のような方法がある。

[0300] (総送信レート決定方法A10)

装置間コネクション終端部からのコネクション毎輻輳情報に依存して決定する。たとえば、各装置間コネクション終端部からのコネクション毎輻輳情報の、ボトルネック物理帯域の推定値をRb、空き帯域の推定値をRaとする。所望の実効レートR_targetが、ボトルネックの空き帯域Raより小さければ、 $R_{total} = R_{target} \cdot \alpha$ とし、 $Ra < R_{target} < Rb$ であれば、 $R_{total} = R_{target} \cdot \beta$ と決める。 α 、 β は、あらかじめ決められた値であり、 $1 < \alpha$ 、 $1 < \beta$ である。

[0301] (総送信レート決定方法B10)

所望の実効レートR_targetが現在中継中のコネクション数に見合ったものになるようR_totalを現在中継中のコネクション数に依存して決める。装置間コネクション終端部からのコネクション毎輻輳情報の、ボトルネック物理帯域の推定値をRb、空き帯域の推定値をRaとする。1コネクション当たりR_oのレートが得られる事を望む場合、中継制御部によって登録されている現在のコネクション数mを元に、 $R_{target} = R_o * m$ と決定する。

[0302] 所望の帯域R_targetが、ボトルネックの空き帯域Raより小さければ、 $R_{total} = R_{target} \cdot \alpha$ とし、 $Ra < R_{target} < Rb$ であれば、 $R_{total} = R_{target} \cdot \beta$ と決める。 α 、 β は、あらかじめ決められた値であり、 $1 < \alpha$ 、 $1 < \beta$ である。

[0303] (総送信レート決定方法C10)

所望の実効レートR_targetが得られるように、装置間コネクション終端部からのコネクション毎輻輳情報をを利用してR_totalを決める。例えば、時間T毎に、総送信レートを更新するものとし、時刻t-Tから時刻tの間に、各中継コネクションがC個のパケットを送り、L回のパケットロスを検知したとすると、パケットロス率は、 $P = L/C$ である。同じ輻輳状況が続くと仮定し、目標レート(R_target)にするために、 $R_{total} = R_{target} / (1 - P)$ とする。ただし、極端な輻輳を避けるため、あるデータロス率の閾値P_0に対して、 $P > P_0$ の場合には、 $R_{total} = R_{target} / (1 - P_0)$ とする。

[0304] (総送信レート決定方法D10)

出来るだけ大きな実効レートが得られるように、R_totalを各装置間コネクション終端

部からのコネクション毎輻輳情報を利用して決める。例えば、時間T毎に、総送信レートを更新するものとし、時刻t-Tから時刻tの間に、各中継コネクションがC個のパケットを送り、L回のパケットロスを検知したとすると、パケットロス率は、 $P=L/C$ である。過去の送信レート R_{total} とその時のデータロス率Pから、送信レートと実際に受信側に届いたレート($=R_{total} * (1-P)$)の関係を求めた表を作る。この表を作るデータは、順次更新してゆくとする。実際に受信側に届いたレートが最大になる送信レートの値を R_{total} とする。ただし、過去の送信レート R_{total} がまだ十分に蓄積されていなければ、あらかじめ設定されている R_{total} を使う。

[0305] (総送信レート決定方法E10)

現在中継しているアプリケーションに見合った所望の実効レート R_{target} が得られるように、 R_{total} をアプリケーション情報解析部からの情報も利用して決める。例えば、ストリーミングコネクションのみを中継している時、アプリケーション情報解析部からストリーミングのレート情報を得て、それらの合計値を R_{target} とし、送信レート決定方法C3に従って R_{total} を求める。

[0306] (総送信レート決定方法F10)

ボトルネックを共有する他のトラヒックに対する優先度付けをした、実効レートが得られるように、 R_{total} を、現在のコネクション数m、および、装置間コネクション終端部からのコネクション毎輻輳情報を利用して決める。例えば、コネクション毎輻輳情報から得られる、中継コネクションが通るネットワークのパケットロス率P、RTTを使ってTCPの1本あたりの実効レートを推定し、その α 倍のレートをm倍してもよい。TCPの1本あたりの実効レートの推定方法としては、M. Handley, S. Floyd, J. Padhye, J. Widmer, "TCP Friendly Rate Control (TFRC): Protocol Specification," Internet Society, RFC 3448, January 2003.の3.1節の"throughput equation"のXで与えられているような計算式でもよい。即ち、中継コネクション終端部からのコネクション毎輻輳情報を利用し、パケットロス発生率p(pの計算の際には、RTT以内のパケットロスも1つのロスイベントと考えて計算する)、RTTの指數平均値R、TCPのタイムアウト時間の推定値 t_{RTO} (例えば、Rの2倍とする)、中継コネクションのあるネットワークのTCPのパケットサイズの平均値をs(バイト)(sは経験的に知られている値でも良い)、 $b=2$ として、

[0307] [数29]

$$X = \frac{s}{R \cdot \sqrt{2 \cdot b \cdot p / 3 + (t_RTO \cdot (3 \sqrt{3 \cdot b \cdot p / 8} \cdot p \cdot (1 + 32 \cdot p^2)))}} \quad (\text{bit/sec})$$

と求め、 $R_{\text{total}} = m \cdot \alpha \cdot X$ (bit/sec)と決定する。 α を1より大きくすれば、1本あたりの送信レートを他のトラヒックに対して高くする事が出来、中継するトラヒックの他のトラヒックに対する優先度を相対的に上げることが出来る。逆に、 α を1未満にすれば、1本あたりの送信レートを他のトラヒックに対して低くする事が出来、中継するトラヒックの他のトラヒックに対する優先度を相対的に下げる事が出来る。

[0308] このようにして決めた R_{total} を各中継コネクションへ割り当てる。このレート割り当ては、現在のコネクション数や、アプリケーション情報解析部からの情報に依存して決めても良く、例えば、次のような方法が考えられる。

[0309] (送信レート割り当て方法1)

全コネクションに公平にレートを割り当てる。即ち、コネクションが m 本あれば、 R_{total}/m のレートを各コネクションに割り当てる。

[0310] (送信レート割り当て方法2)

コネクション識別子のポート番号情報に依存して、決められたルールに従って割り当てる。たとえば、ポート番号毎に優先度を付け、その優先度に応じてレートを割り当てる。

[0311] (送信レート割り当て方法3)

アプリケーション情報解析部からの情報に依存して決定する。例えば、HTTPコネクションでデータ転送を行うアプリケーションの複数のコネクションを中継している場合に、アプリケーション情報解析部から得られた、HTTPのContent-typeフィールドの値に依存して優先度を付けた帯域割り当てを行う。別の例としては、HTTPでファイル取得を行うアプリケーションの複数のコネクションを中継している場合に、アプリケーション情報解析部から得られたHTTPのContent-lengthフィールドの情報に依存して、短いサイズのデータを転送しているコネクションにより高い帯域を割り当てる。また別の例として、Ns本のストリーミングタイプのアプリケーションのコネクションと、Nf本のHTTPを使ったファイル転送タイプのアプリケーションを中継している場合を考える。

アプリケーション情報解析部から得られた、各ストリーミングのレートが $RS_i(i=1, \dots, N_s)$ とし、アプリケーション情報解析部から得られた、ファイル転送タイプのコネクションのファイルサイズを $FL_i(i=1, \dots, N_f)$ とする。ストリーミングタイプのコネクションについては、公平なレート(R_{total} を現在のコネクション数 $N_s + N_f$ で割ったもの)より低いレートのものには、そのストリーミングのレート $RS_i(i=1, \dots, N_s)$ を割り当てる。残ったレート

[0312] [数30]

$$(R_{total} - \sum_{i=1}^{N_s} RS_i)$$

を、まだ割り当てをしていないストリーミングコネクションに優先的に割り当てる、さらに残ったレートを、ファイル転送タイプのコネクションのうちファイルサイズが短いもの順に高いレートを割り当てる。

[0313] [動作の説明]

次に、本発明の第10の実施の形態の動作は、本発明の第3の実施の形態の動作と同様である。違いは、装置間コネクション終端部が、その内部のインライン計測部からの、ボトルネックの物理帯域、空き帯域の推定値の情報を付与して、送信レート制御部へ、コネクション毎輻輳情報を通知する事と、送信レート制御部が、送信レート決定の際に、A10、B10に示したような、インライン計測部からの、ボトルネックの物理帯域、空き帯域の推定値の情報も用いた送信レートの決定を行う点のみである。

[0314] [発明の第10の実施の形態の効果]

発明の第10の実施の形態では、本発明の第3の実施の形態の効果に加え、次のような効果が得られる。送信レート制御部において、総送信レート決定方法A10、B10の方法を用いることにより、ボトルネックの物理帯域や空き帯域の推定値も考慮した、送信レート決定を行うことにより、高い精度での所望の実効レートの確保や、他のトラヒックに対して高い精度で優先度をつけた実効レートを確保する事が可能になることがある。

請求の範囲

[1] 複数のトランスポート層コネクションをそれぞれトランスポート層で終端し、前記各トランスポート層コネクション上のデータフローをそれぞれ別のトランスポート層コネクションへ中継するトランスポート層中継方法であって、
中継に使用中の中継コネクションの総送信レートを決定し、前記総送信レートを分割して、中継に使用中の中継コネクションの送信レートとして割り当てるトランスポート層中継方法。

[2] 請求項1記載のトランスポート層中継方法であって、
前記総送信レートを、中継しているトランスポート層コネクションの数、および、中継コネクションが経由するネットワークの輻輳状況に応じて決定するトランスポート層中継方法。

[3] 請求項1記載のトランスポート層中継方法であって、
中継に使用中の中継コネクションの全体の実効レートが所望のレートになるように、前記総送信レートを、中継しているトランスポート層コネクションの数、および、中継コネクションが経由するネットワークの輻輳状況に応じて決定するトランスポート層中継方法。

[4] 請求項1記載のトランスポート層中継方法において、
中継に使用中の中継コネクションの全体の実効レートが、中継コネクションとボトルネックを共有する中継コネクション以外のトラヒックに対して優先度付け出来るように、前記総送信レートを、中継しているトランスポート層コネクションの数、および、中継コネクションが経由するネットワークの輻輳状況に応じて決定するトランスポート層中継方法。

[5] 請求項1記載のトランスポート層中継方法において、
前記総送信レートを、前記各中継コネクション上の前記データフロー内のアプリケーション情報に依存して、前記各中継コネクションの送信レートに分割して割り当てるトランスポート層中継方法。

[6] 請求項1記載のトランスポート層中継方法において、
中継コネクションが経由するネットワークの輻輳状況を計測パケットによって推定し

た結果も利用して、前記総送信レートを決定するトランスポート層中継方法。

[7] 請求項1記載のトランスポート層中継方法において、
中継コネクションが経由するネットワークの輻輳状況を、中継するパケットによって推定した結果も利用して、前記総送信レートを決定するトランスポート層中継方法。

[8] 複数の端末とのトランスポート層コネクションを、それぞれトランスポート層で終端する複数の端末側コネクション終端部と、トランスポート層中継装置間の複数のトランスポート層コネクションをそれぞれ終端する複数の装置間コネクション終端部を持ち、前記各端末側コネクション終端部と前記各装置間コネクション終端部の間でトランスポート層データを中継する、トランスポート層中継装置において、
前記装置間コネクション終端部は、送信レート制御部から通知される送信レートに従って送信を行い、
送信レート制御部は、中継に使用中の装置間コネクション終端部全体の総送信レートを決定し、前記総送信レート分割し、分割して割り当てたそれぞれのレートを前記中継に使用中の各装置間コネクション終端部に通知するトランスポート層中継装置。

[9] 請求項8記載のトランスポート層中継装置において、
前記送信レート制御部が、前記総送信レートを、中継しているトランスポート層コネクションの数、および、各装置間コネクション終端部から通知されるコネクション毎の輻輳情報に応じて決定するトランスポート層中継装置。

[10] 請求項8記載のトランスポート層中継装置において、
前記送信レート制御部が、中継に使用中の中継コネクションの全体の実効レートが所望のレートになるように、前記総送信レートを、中継しているトランスポート層コネクションの数、および、各装置間コネクション終端部から通知されるコネクション毎の輻輳情報に応じて決定するトランスポート層中継装置。

[11] 請求項8記載のトランスポート層中継装置において、
前記送信レート制御部が、中継に使用中の中継コネクションの全体の実効レートが、中継コネクションとボトルネックを共有する中継コネクション以外のトラヒックに対して優先度付け出来るように、前記総送信レートを、中継しているトランスポート層コネクションの数、および、各装置間コネクション終端部から通知されるコネクション毎の輻

輶情報に応じて決定するトランSPORT層中継装置。

[12] 請求項8記載のトランSPORT層中継装置において、
前記各端末側コネクション終端部と前記各装置間コネクション終端部の間でトランSPORT層データを中継する時に、前記トランSPORT層データ内のアプリケーション情報を探するアプリケーション情報解析部を持ち、且つ、
前記レート制御部が、前記総送信レートを、前記アプリケーション情報解析部からのアプリケーション情報に基づいて、前記各中継コネクションの送信レートに分割して割り当てるトランSPORT層中継装置。

[13] 請求項8記載のトランSPORT層中継装置において、
中継コネクションが経由するネットワークの輶状況を、計測パケットによって推定するネットワーク状況推定部を持ち、且つ、
前記送信レート制御部が、前記ネットワーク状況推定部が推定した結果も利用して、前記総送信レートを決定するトランSPORT層中継装置。

[14] 請求項8記載の何れか1つのトランSPORT層中継装置において、
中継コネクションが経由するネットワークの輶状況を、中継するパケットによって推定するインライン計測部を持ち、且つ、
前記送信レート制御部が、前記インライン計測部が推定した結果も利用して、前記総送信レートを決定するトランSPORT層中継装置。

[15] 複数のトランSPORT層コネクションをそれぞれトランSPORT層で終端し、前記各トランSPORT層コネクション上のデータフローを1つのトランSPORT層コネクションへ集約して中継するトランSPORT層中継方法において、
中継コネクションの総送信レートを決定し、前記総送信レートを分割したレートに従って、前記の各トランSPORT層コネクションからのデータフローを中継コネクションへ集約するトランSPORT層中継方法。

[16] 請求項15記載のトランSPORT層中継方法において、
前記総送信レートを、中継しているトランSPORT層コネクションの数、および、中継コネクションが経由するネットワークの輶状況に応じて決定するトランSPORT層中継方法。

[17] 請求項15記載のトランSPORT層中継方法において、
中継に使用中の中継コネクションの全体の実効レートが所望のレートになるように、
前記総送信レートを、中継しているトランSPORT層コネクションの数、および、中継コ
ネクションが経由するネットワークの輻輳状況に応じて決定するトランSPORT層中継
方法。

[18] 請求項15記載のトランSPORT層中継方法において、
中継に使用中の中継コネクションの全体の実効レートが、中継コネクションとボトル
ネックを共有する中継コネクション以外のトラヒックに対して優先度付け出来るように、
前記総送信レートを、中継しているトランSPORT層コネクションの数、および、中継コ
ネクションが経由するネットワークの輻輳状況に応じて決定するトランSPORT層中継
方法。

[19] 請求項15記載のトランSPORT層中継方法において、
前記総送信レートを、前記データフロー内のアプリケーション情報に依存して分割
したレートに従って、前記各トランSPORT層コネクションからのデータフローを中継コ
ネクションへ集約するトランSPORT層中継方法。

[20] 請求項15記載のトランSPORT層中継方法において、
中継コネクションが経由するネットワークの輻輳状況を計測パケットによって推定し
た結果も利用して、前記総送信レートを決定するトランSPORT層中継方法。

[21] 請求項15記載のトランSPORT層中継方法において、
中継コネクションが経由するネットワークの輻輳状況を中継するパケットによって推
定した結果も利用して、前記総送信レートを決定するトランSPORT層中継方法。

[22] 端末とのトランSPORT層コネクションをトランSPORT層で終端する複数の端末側コ
ネクション終端部と、トランSPORT層中継装置間のトランSPORT層コネクションを終
端する1つの装置間コネクション終端部を持ち、前記各端末側コネクション終端部か
らのトランSPORT層データを1つに集約して前記装置間コネクション終端部へ受け渡
すMUX-DEMUX部を持つトランSPORT層中継装置において、
前記装置間コネクション終端部は、送信レート制御部から通知される総送信レート
に従って送信を行い、

前記MUX-DEMUX部は、送信レート制御部から通知されるレートの分配に従って端末側コネクション終端部からのデータを集約し、

送信レート制御部は、前記装置間コネクション終端部の総送信レートを決定して通知すると共に、前記総送信レートを分割したレートの分配を前記MUX-DEMUX部へ通知するトランスポート層中継装置。

[23] 請求項22記載のトランスポート層中継装置において、

前記送信レート制御部が、前記総送信レートを、中継しているトランスポート層コネクションの数、および、装置間コネクション終端部から通知されるコネクションの輻輳情報に応じて決定するトランスポート層中継装置。

[24] 請求項22記載のトランスポート層中継装置において、

前記送信レート制御部が、中継コネクションの実効レートが所望のレートになるよう前記総送信レートを、中継しているトランスポート層コネクションの数、および、装置間コネクション終端部から通知されるコネクションの輻輳情報に応じて決定するトランスポート層中継装置。

[25] 請求項22記載のトランスポート層中継装置において、

前記送信レート制御部が、中継コネクションの実効レートが中継コネクションとボトルネックを共有する中継コネクション以外のトラヒックに対して優先度付け出来るように、前記総送信レートを、中継しているトランスポート層コネクションの数、および、各装置間コネクション終端部から通知されるコネクションの輻輳情報に応じて決定するトランスポート層中継装置。

[26] 請求項22記載のトランスポート層中継装置において、

前記各端末側コネクション終端部と前記MUX-DEMUX部の間でトランスポート層データを受け渡す時にトランスポート層データ内のアプリケーション情報を解析するアプリケーション情報解析部を持ち、

前記レート制御部が前記アプリケーション情報解析部からのアプリケーション情報に基づいて、前記総送信レートを分割し、前記MUX-DEMUX部へ伝えるレートの分配を決定するトランスポート層中継装置。

[27] 請求項22記載のトランスポート層中継装置において、

中継コネクションが経由するネットワークの輻輳状況を、計測パケットによって推定するネットワーク状況推定部を持ち、且つ、

前記送信レート制御部が、前記ネットワーク状況推定部が推定した結果も利用して、前記総送信レートを決定するトранスポート層中継装置

[28] 請求項22記載のトランスポート層中継装置において、

中継コネクションが経由するネットワークの輻輳状況を、中継するパケットによって推定するインライン計測部を備え、且つ、

前記送信レート制御部が、前記インライン計測部が推定した結果も利用して前記総送信レートを決定するトランスポート層中継装置。

[29] 請求項1記載のトランスポート層中継方法において、

新たなトランスポート層コネクション開設時に、前記総送信レートを決定し、前記総送信レートを分割して各中継コネクションの送信レートとして割り当て、前記新たなトランスポート層プロトコルの開設相手に、割り当てたレートを通知するトランスポート層中継方法。

[30] 請求項8記載のトランスポート層中継装置において、

端末とのトランスポート層コネクション開設時に、前記送信レート制御部から通知される、初期送信レート情報を、端末へ通知するトランスポート層中継装置。

[31] 請求項15記載のトランスポート層中継方法において、

新たなトランスポート層コネクション開設時に、前記総送信レートを決定し、前記総送信レートを分割したレートを、前記の新たなトランスポート層プロトコルの開設相手通知するトランスポート層中継方法。

[32] 請求項22記載のトランスポート層中継装置において、

端末とのトランスポート層コネクション開設時に、前記送信レート制御部から通知される、初期送信レート情報を、端末へ通知するトランスポート層中継装置。

[33] コンピュータを、

複数の端末とのトランスポート層コネクションを、それぞれトランスポート層で終端する複数の端末側コネクション終端部、

前記各端末側コネクション終端部との間でトランスポート層データを中継すると共に

、トランスポート層中継装置間の複数のトランスポート層コネクションをそれぞれ終端し、送信時には送信レート制御部から通知される送信レートに従って送信を行う複数の装置間コネクション終端部、

中継に使用中の装置間コネクション終端部全体の総送信レートを決定し、前記総送信レート分割し、分割して割り当てたそれぞれのレートを前記中継に使用中の各装置間コネクション終端部に通知する送信レート制御部として機能させるためのプログラム。

[34] コンピュータを、

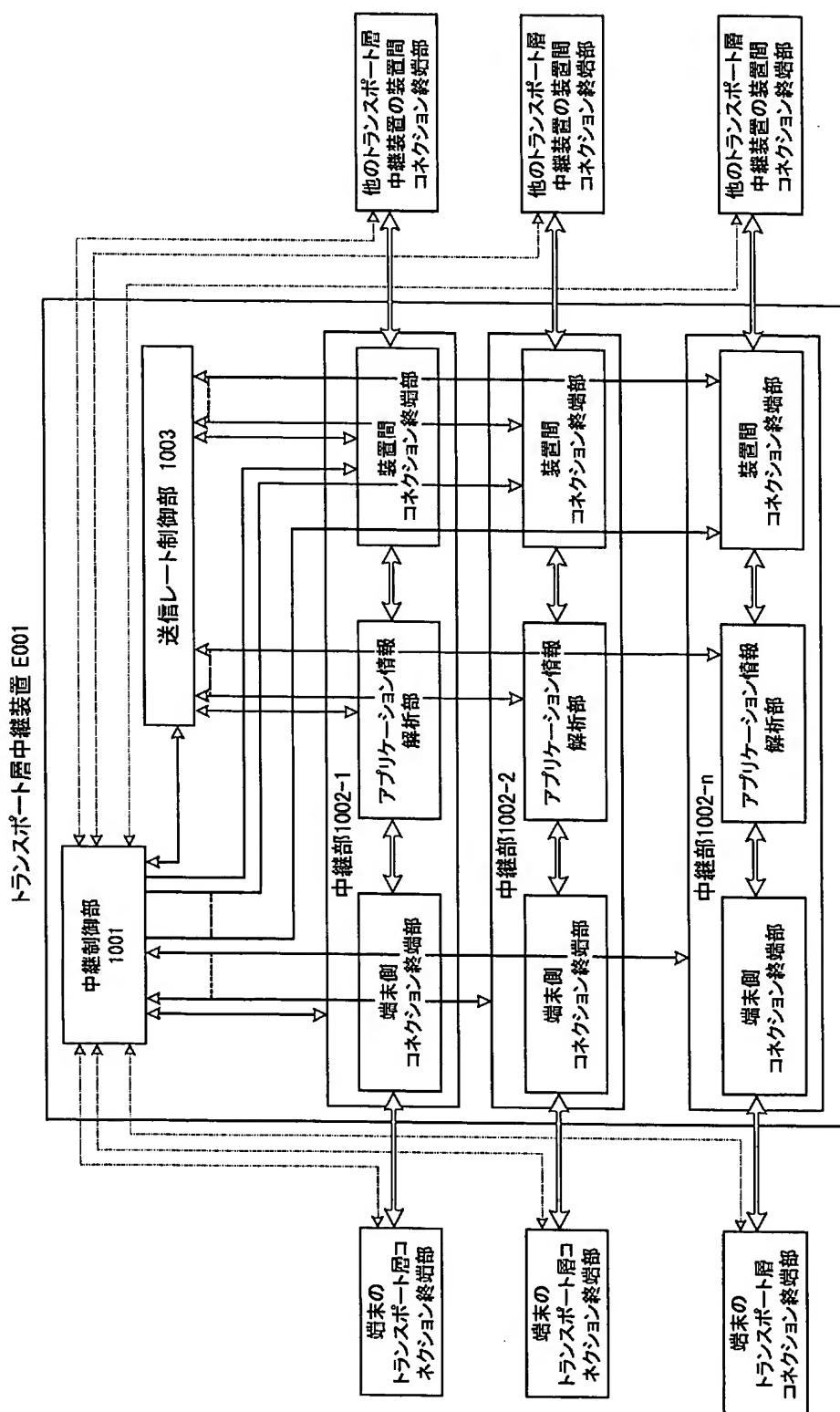
端末とのトランスポート層コネクションをトランスポート層で終端する複数の端末側コネクション終端部、

トランスポート層中継装置間のトランスポート層コネクションを終端し、送信時には送信レート制御部から通知される送信レートに従って送信を行う1つの装置間コネクション終端部、

前記各端末側コネクション終端部からのトランスポート層データを、送信レート制御部から通知されるレートの分配に従って1つに集約して前記装置間コネクション終端部へ受け渡すMUX-DEMUX部、

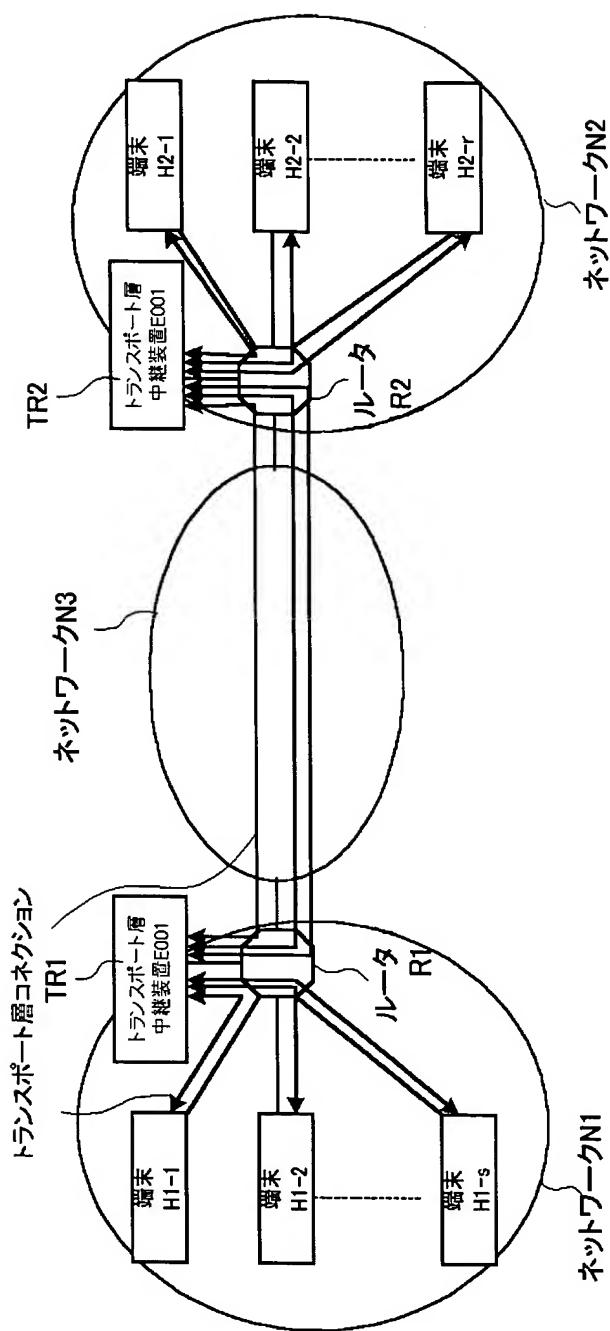
前記装置間コネクション終端部の総送信レートを決定して通知すると共に、前記総送信レートを分割したレートの分配をMUX-DEMUX部へ通知する送信レート制御部として機能させるためのプログラム。

[図1]



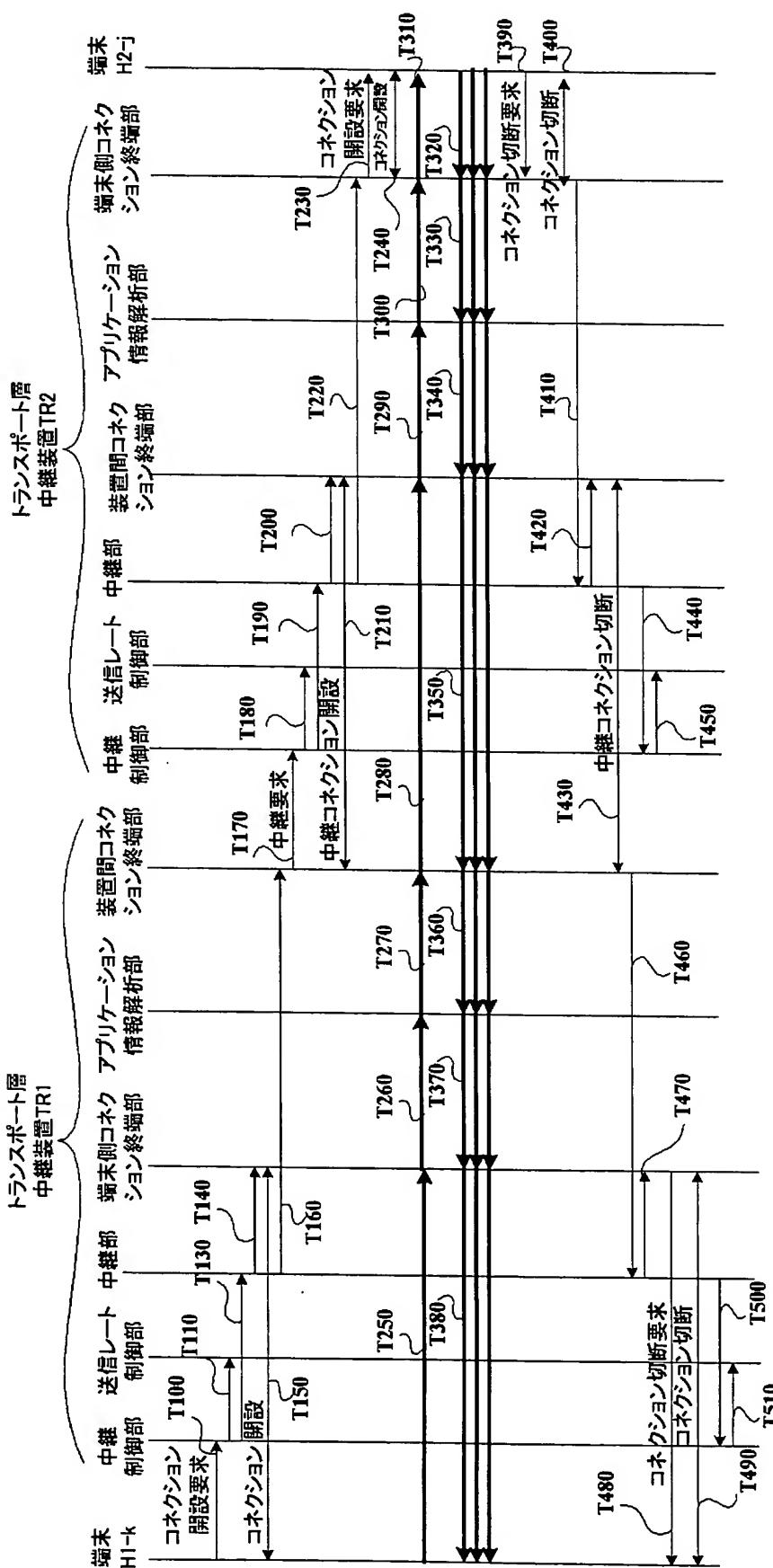
第1の実施の形態のトランスポート層中継装置の構成図

[図2]



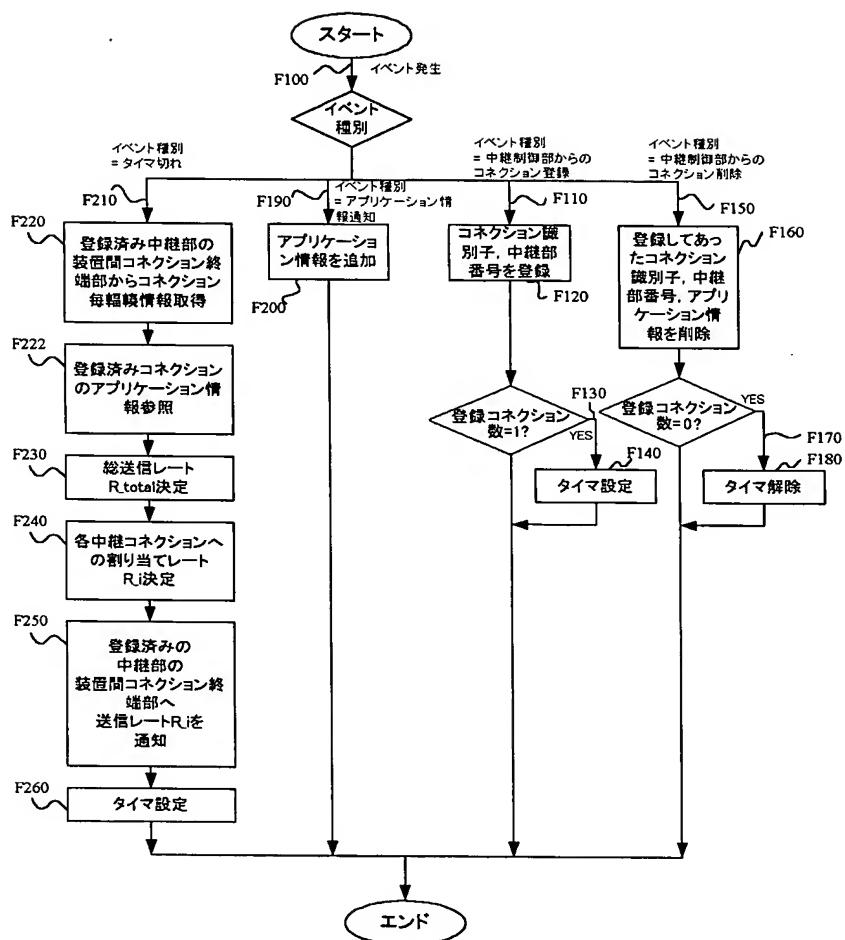
第1の実施の形態のトランSPORT層中継装置を用いたネットワーク構成図

[图3]



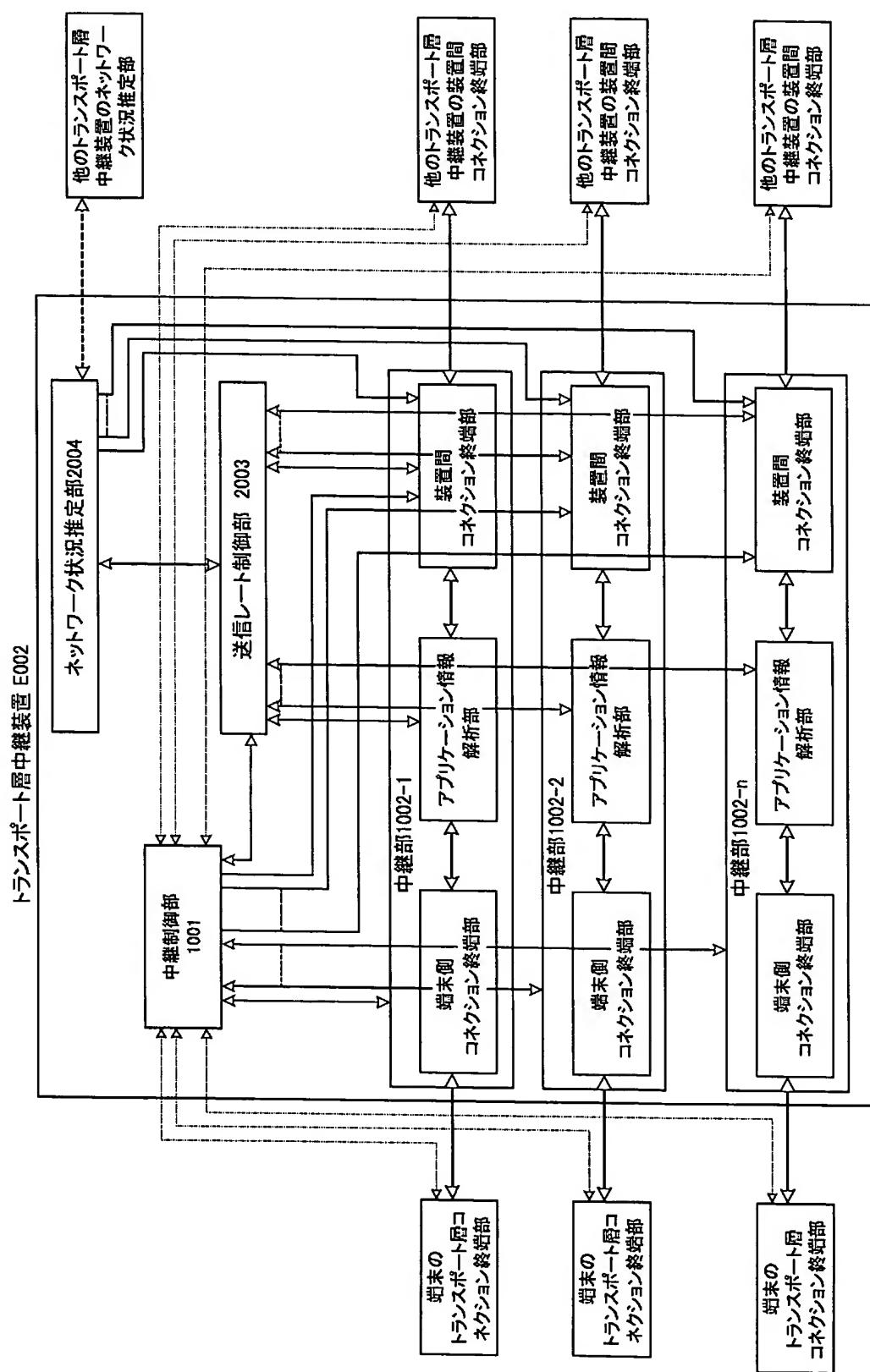
第1の実施の形態の動作説明タイムチャート

[図4]



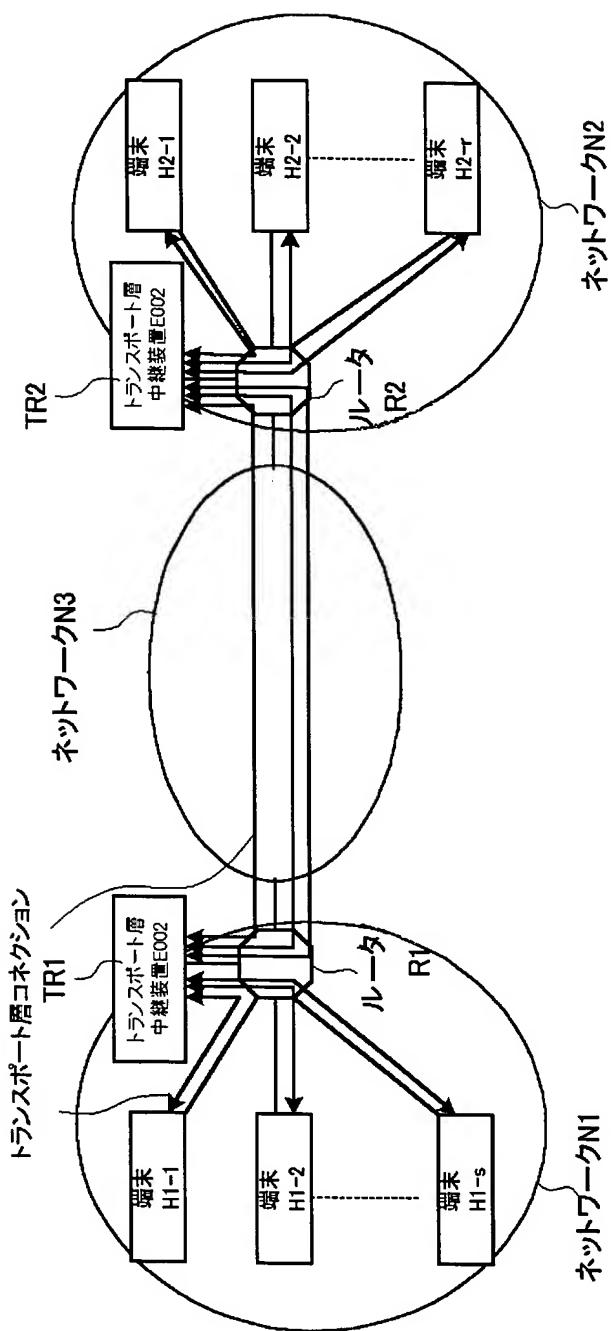
第1の実施の形態のレート制御部の動作フローチャート

[図5]



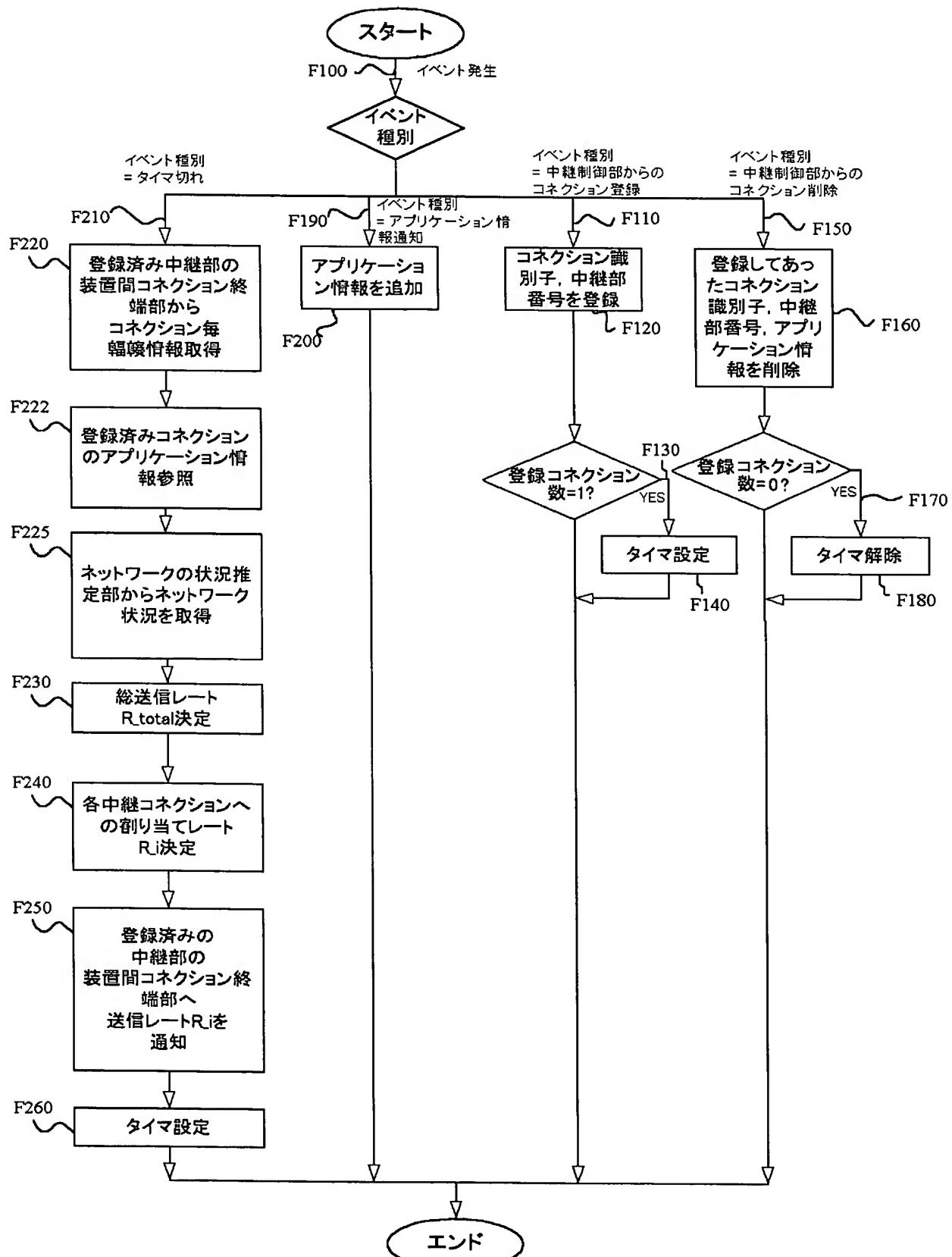
第2の実施の形態のトランSPORT層中継装置の構成図

[図6]



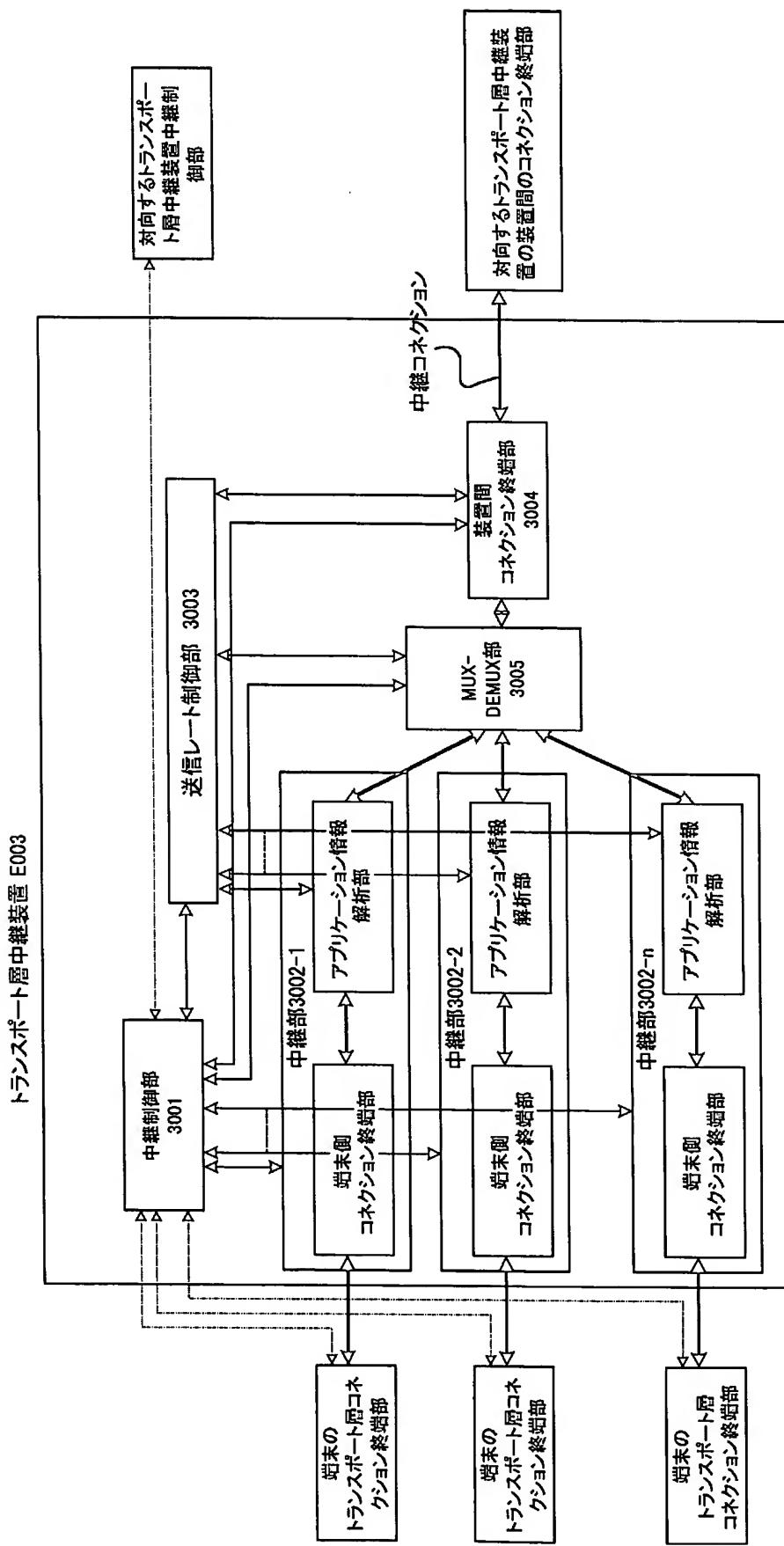
第2の実施の形態のトランスポート層中継装置を用いたネットワーク構成図

[図7]



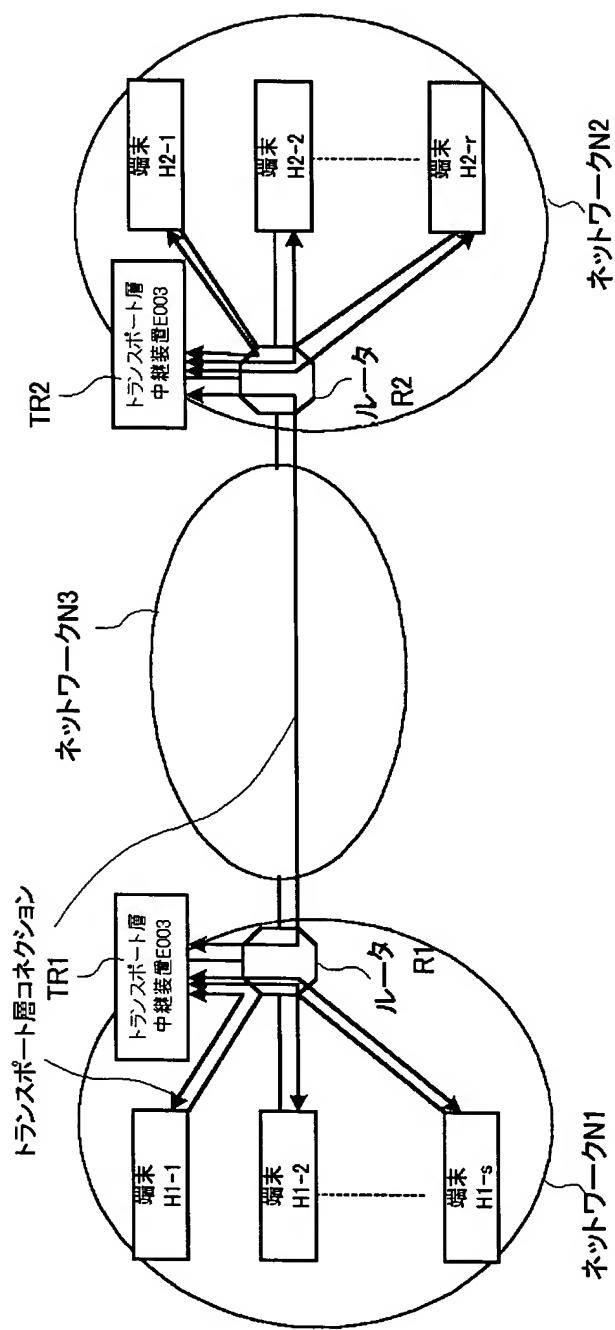
第2の実施の形態のレート制御部の動作フローチャート

[図8]



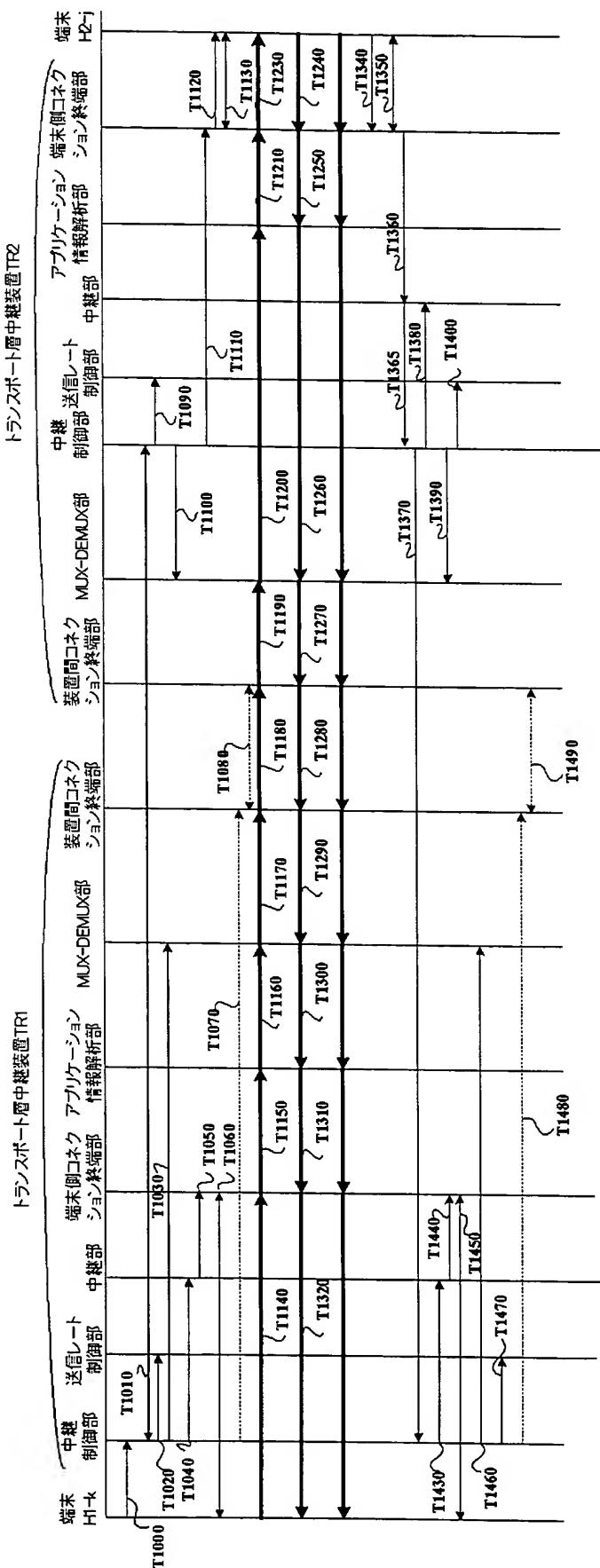
第3の実施の形態のトランスポート層中継装置の構成図

[図9]



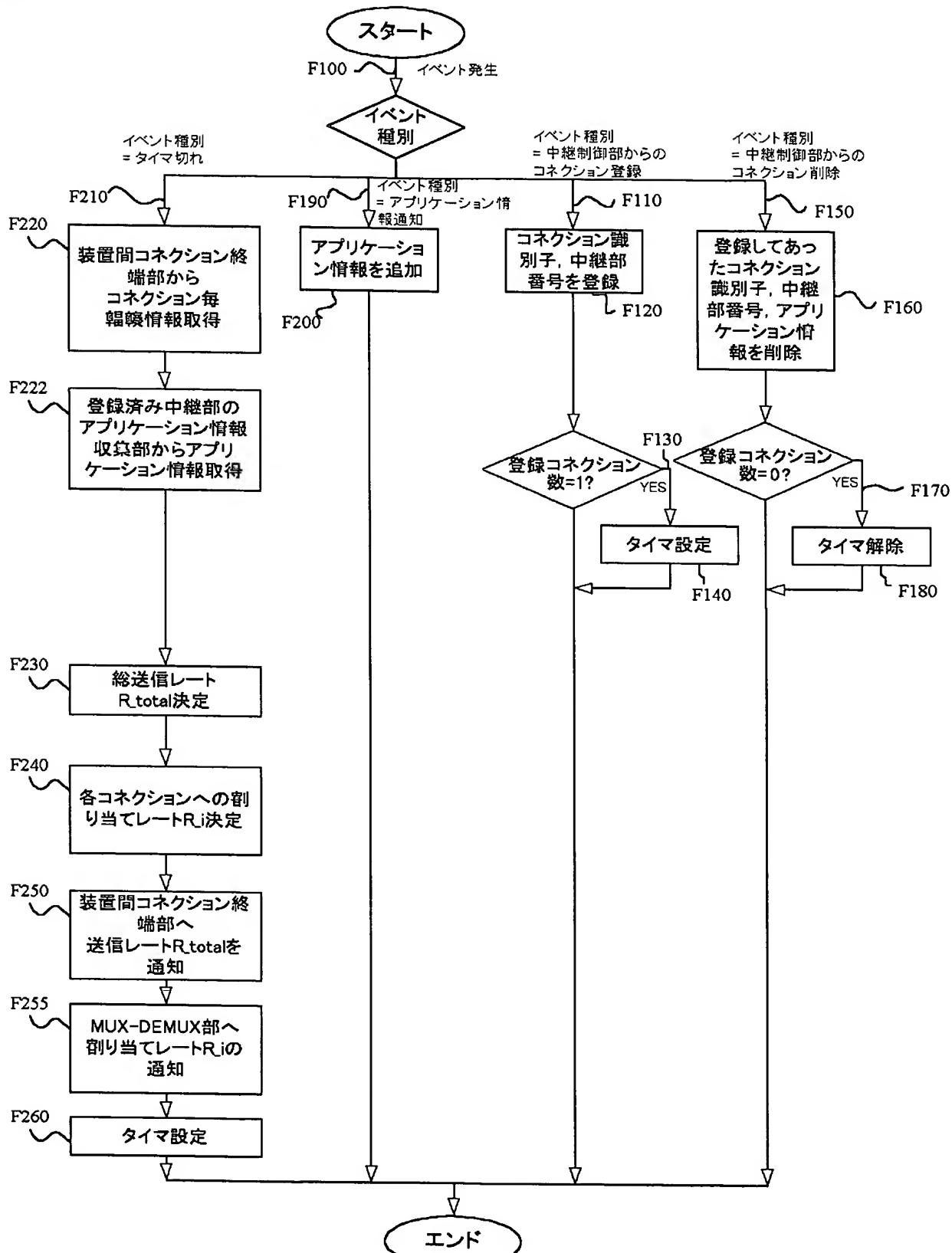
第3の実施の形態のトランスポート層中継装置を用いたネットワーク構成図

[図10]



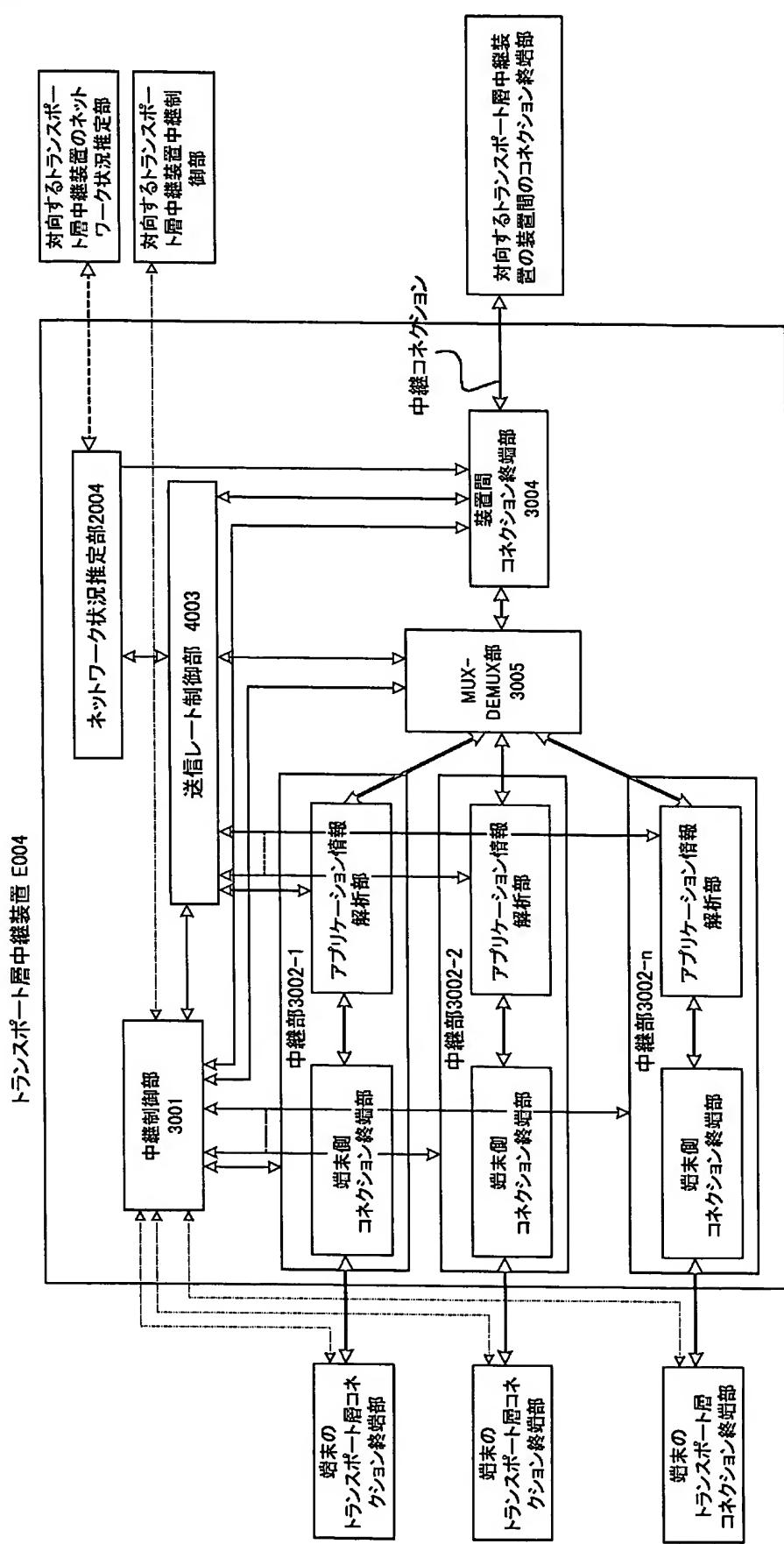
第3の実施の形態の動作説明タイムチャート

[図11]



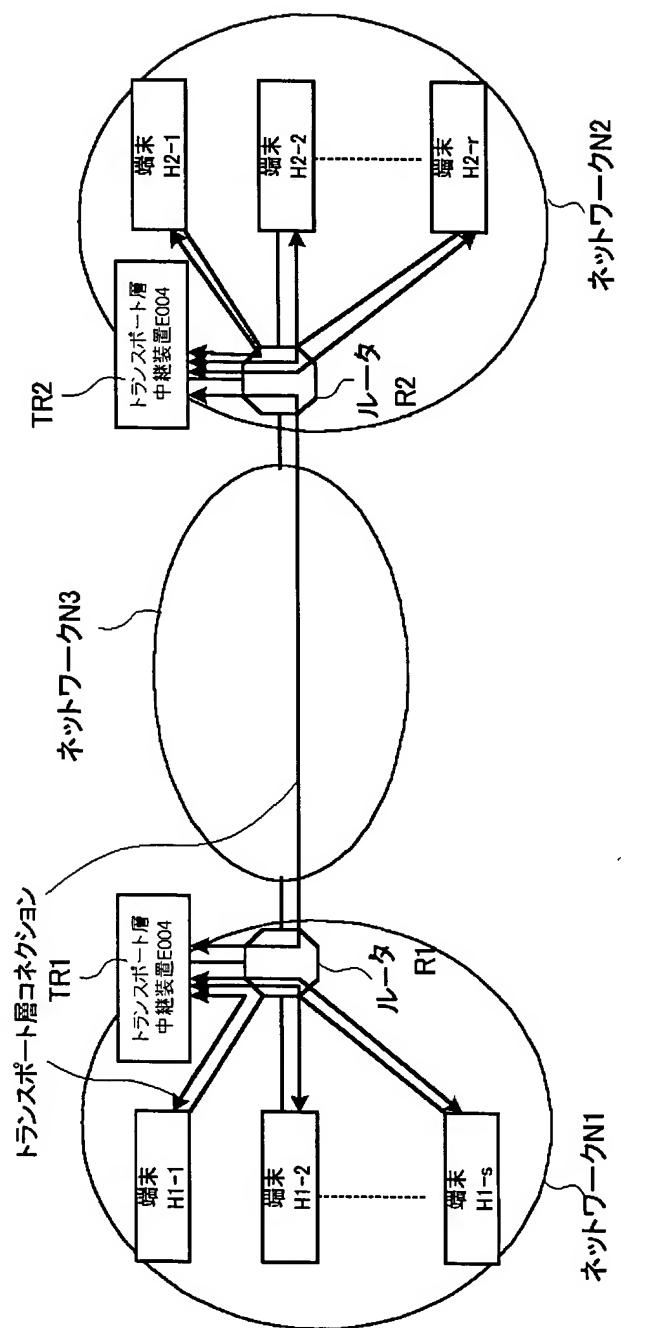
第3の実施の形態のレート制御部の動作フロー チャート

[図12]



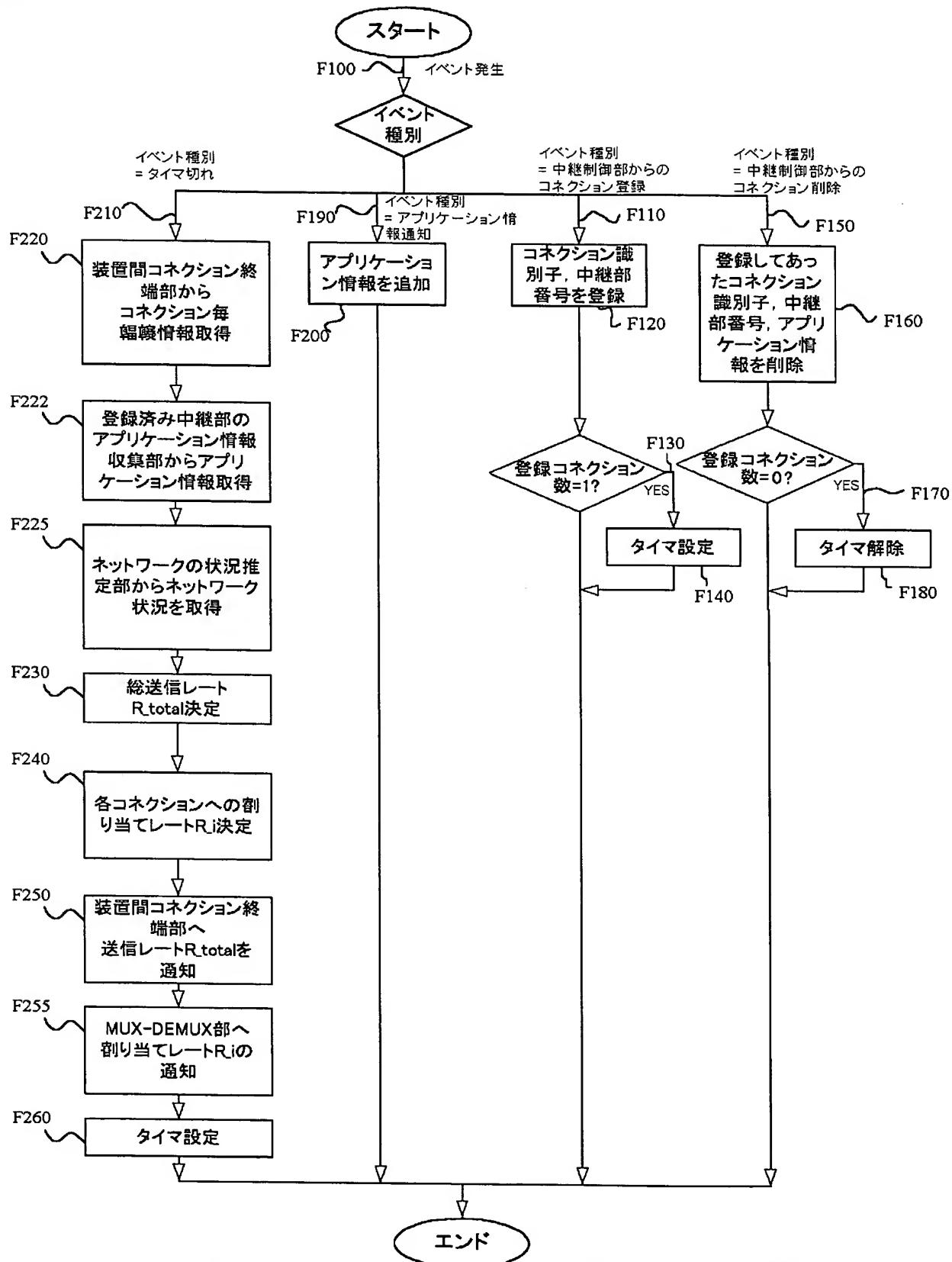
第4の実施の形態のトランスポート層中継装置の構成図

[図13]



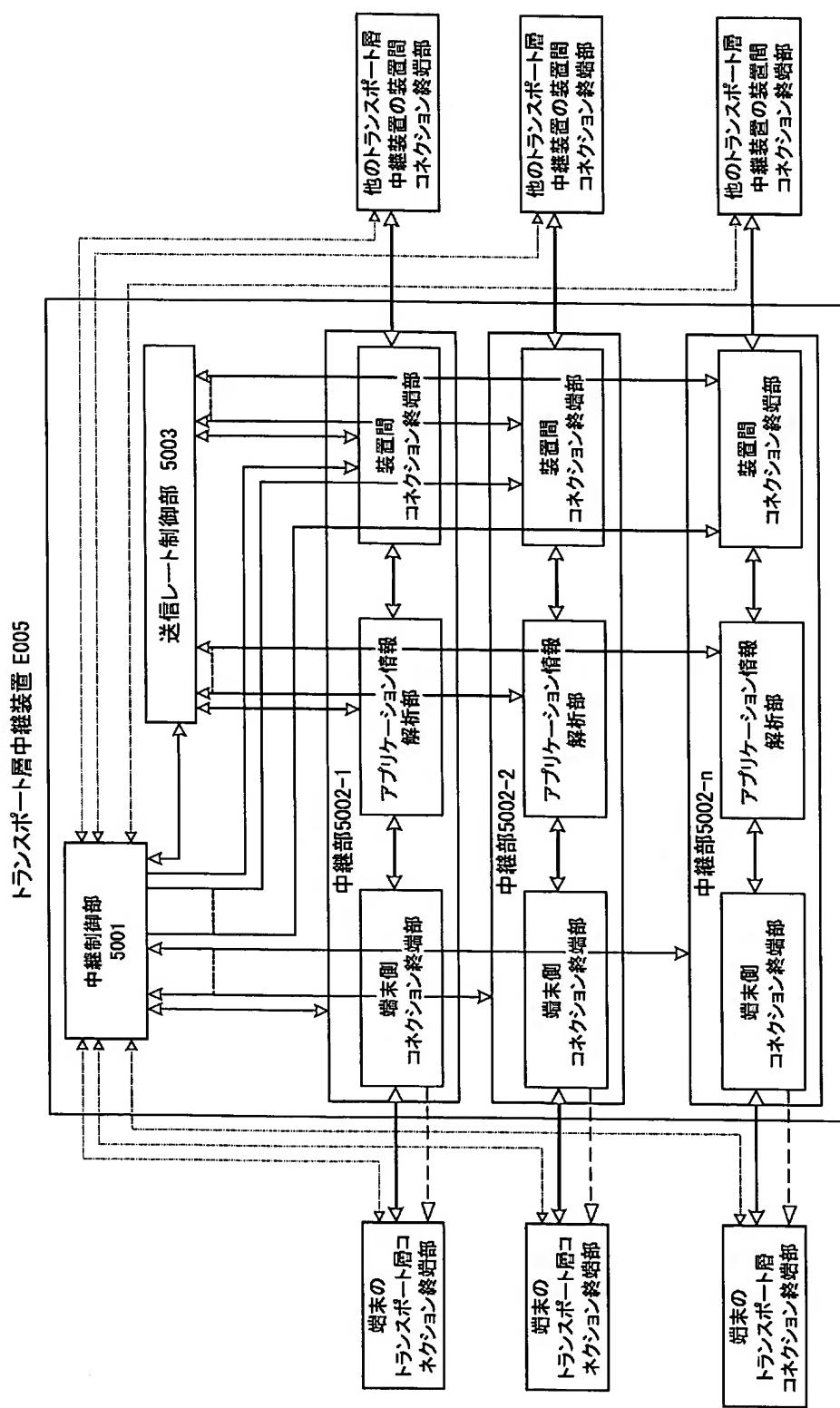
第4の実施の形態のトランSPORT層中継装置を用いたネットワーク構成図

[図14]



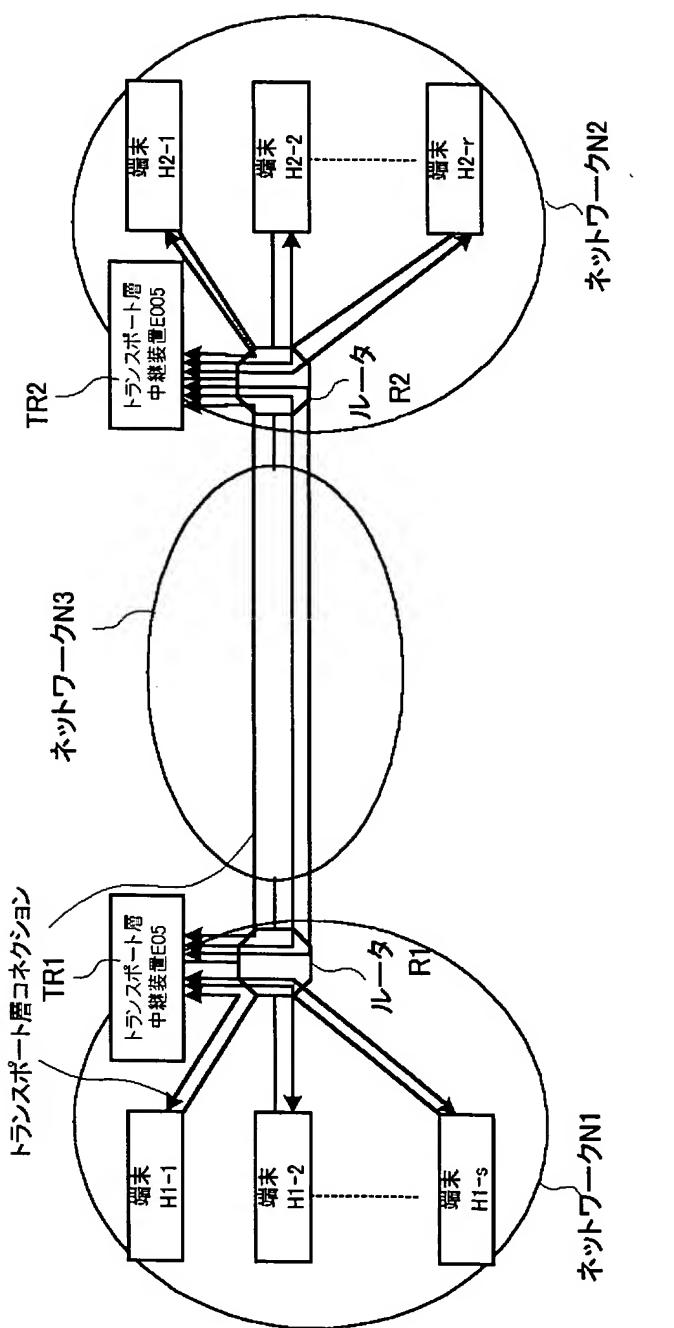
第4の実施の形態のレート制御部の動作フローチャート

[図15]



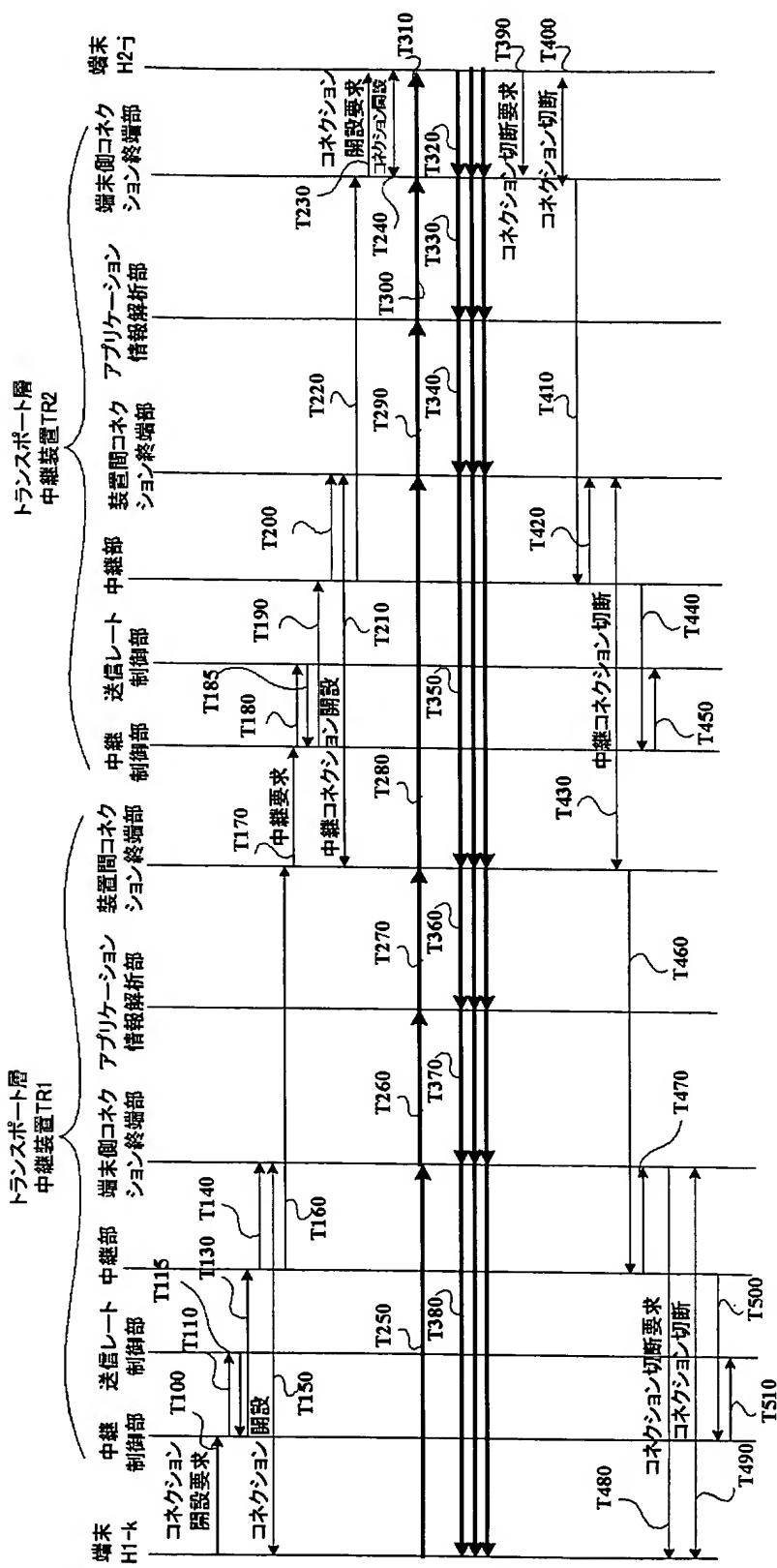
第5の実施の形態のトランスポート層中継装置の構成図

[図16]



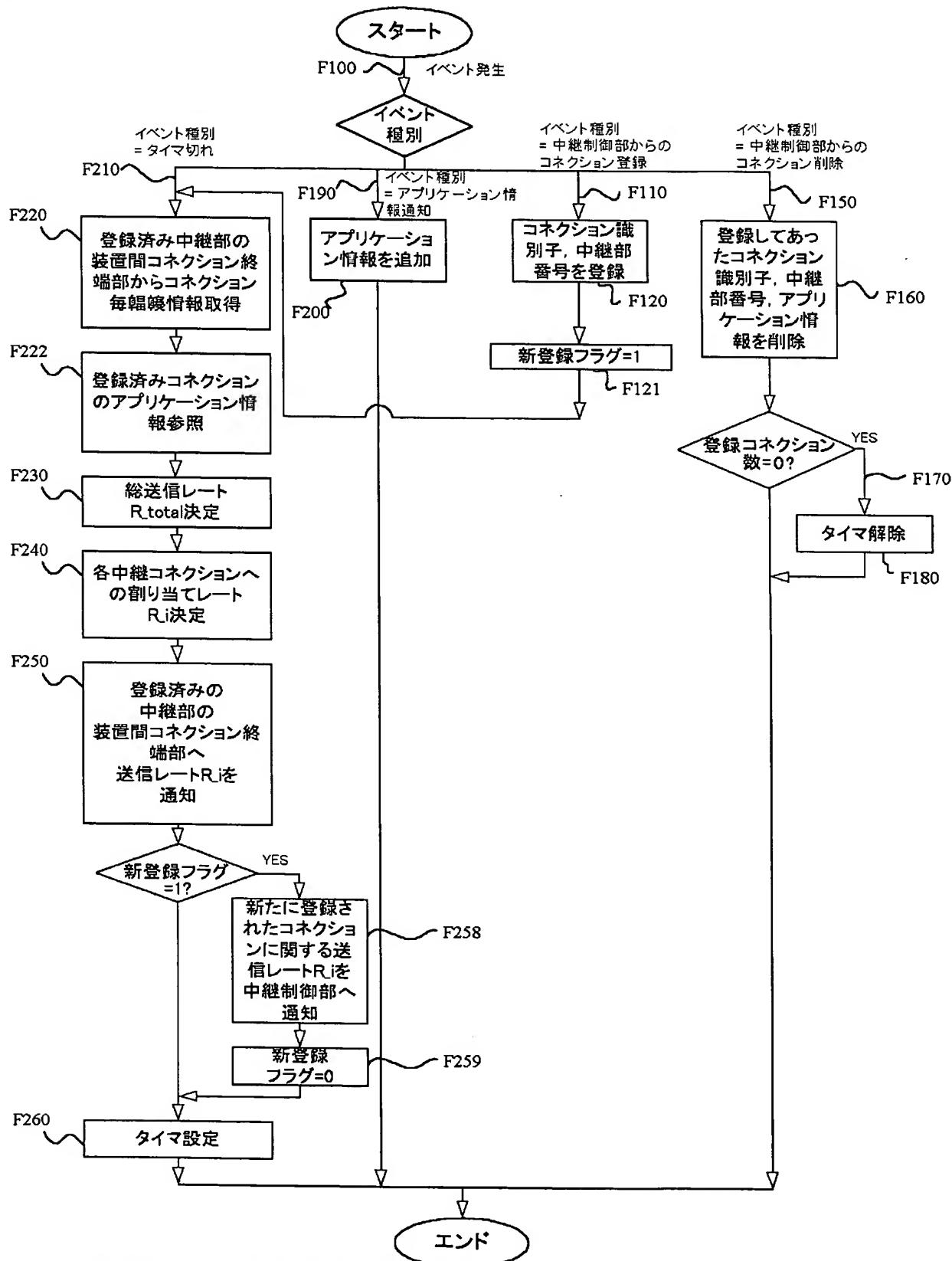
第5の実施の形態のトランスポート層中継装置を用いたネットワーク構成図

[図17]



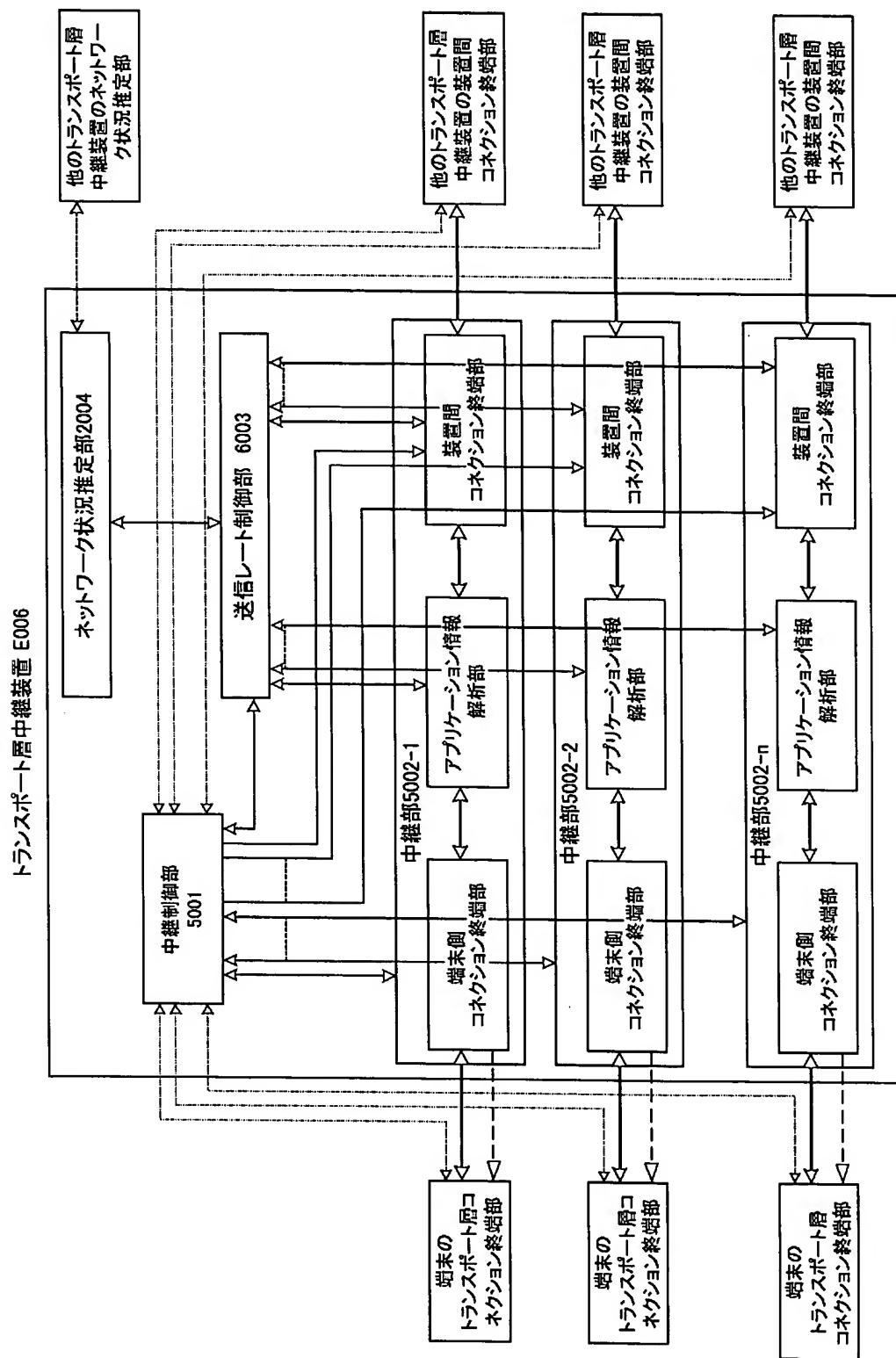
第5の実施の形態の動作説明タイムチャート

[図18]



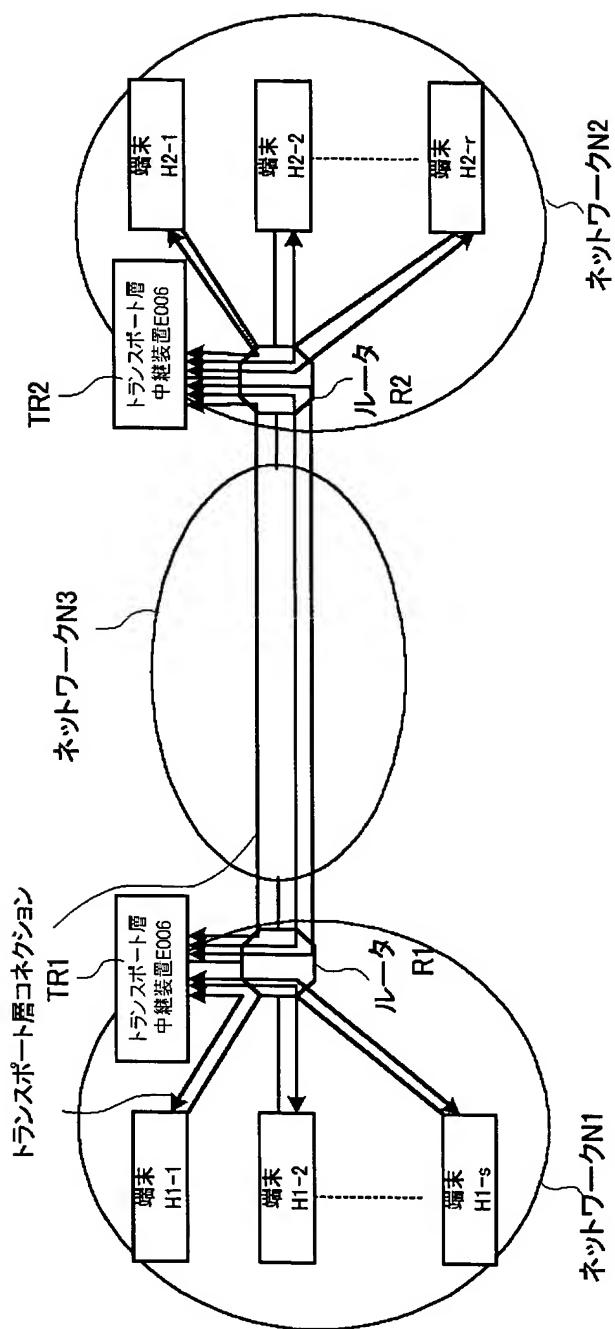
第5の実施の形態のレート制御部の動作フローチャート

[図19]



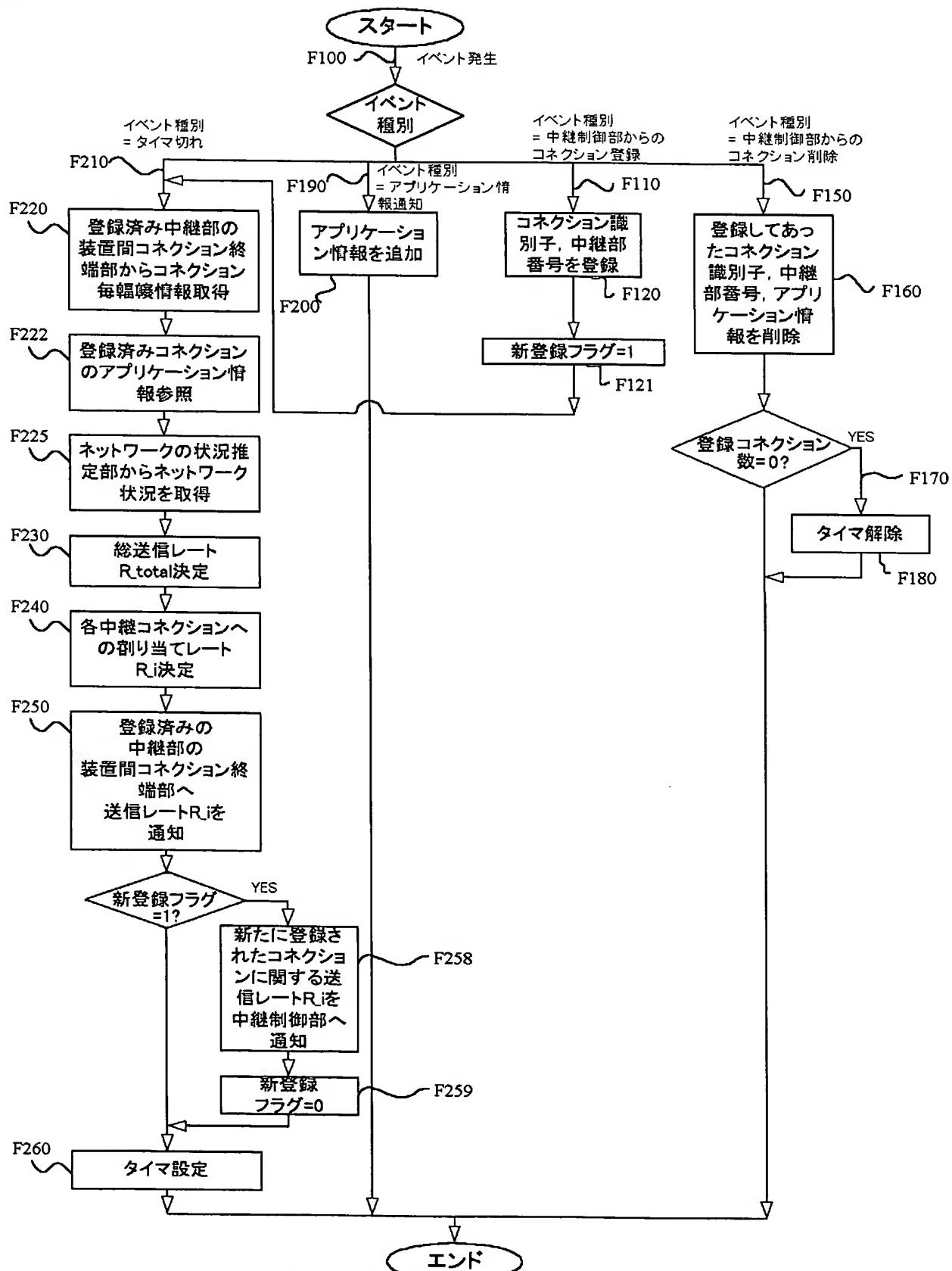
第6の実施の形態のトランスポート層中継装置の構成図

[図20]



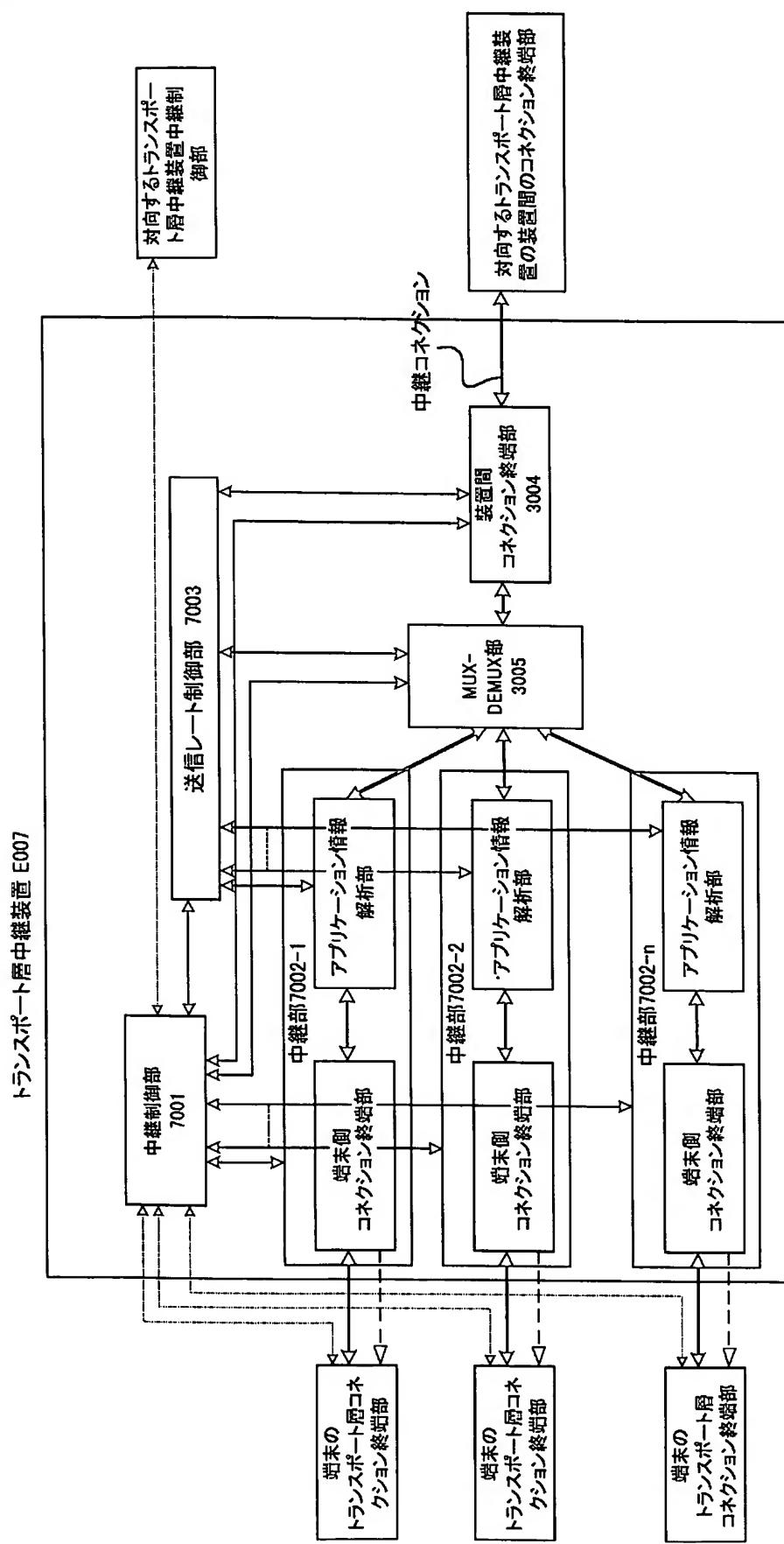
第6の実施の形態のトランスポート層中継装置を用いたネットワーク構成図

[図21]



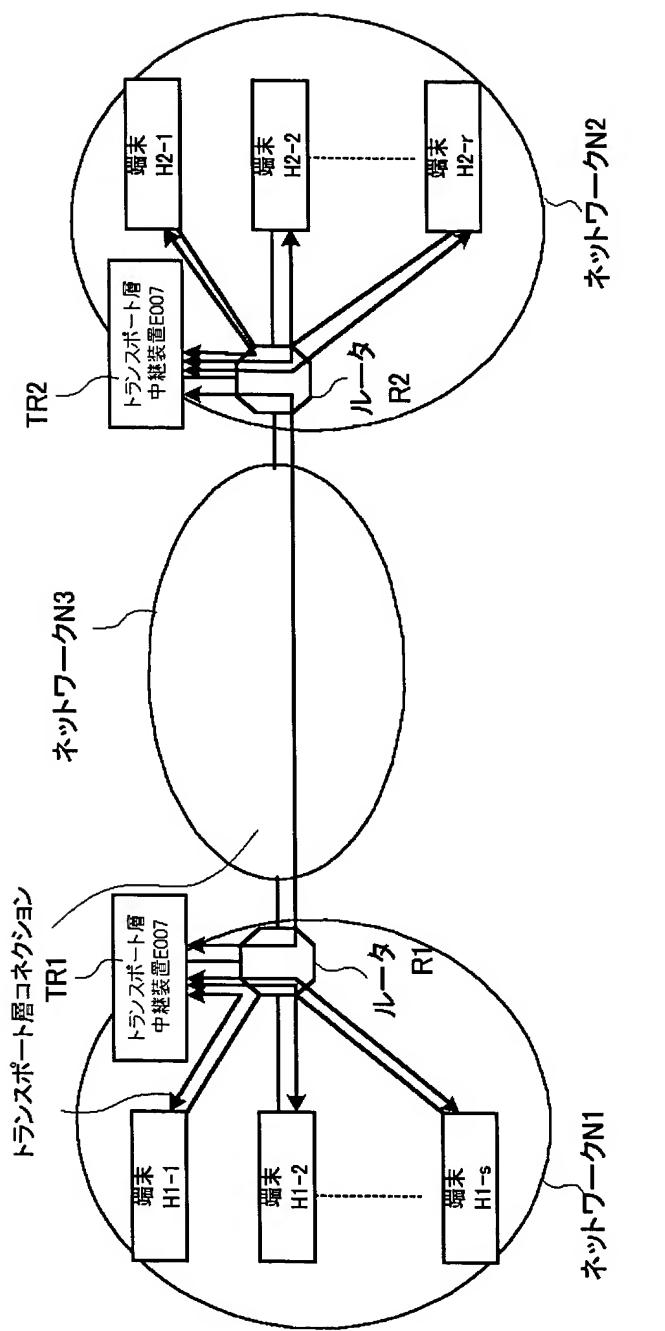
第6の実施の形態のレート制御部の動作フローチャート

[図22]



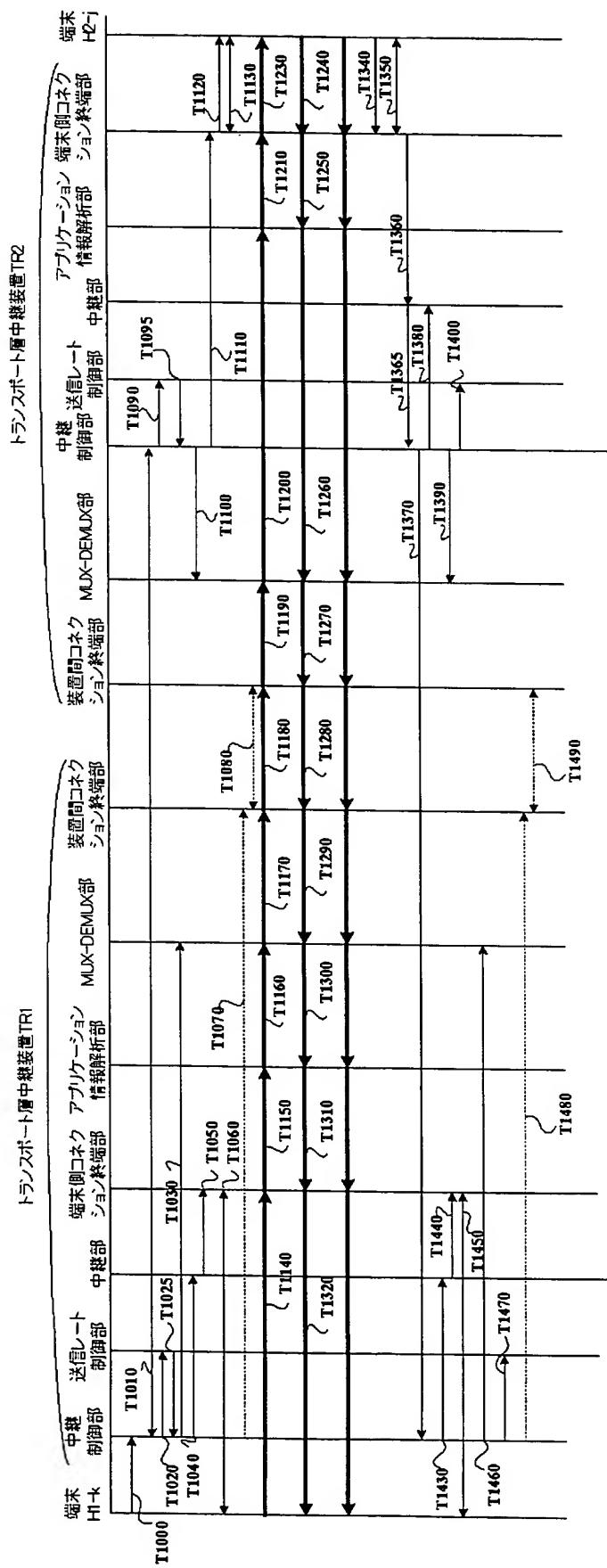
第7の実施の形態のトランスポート層中継装置の構成図

[図23]



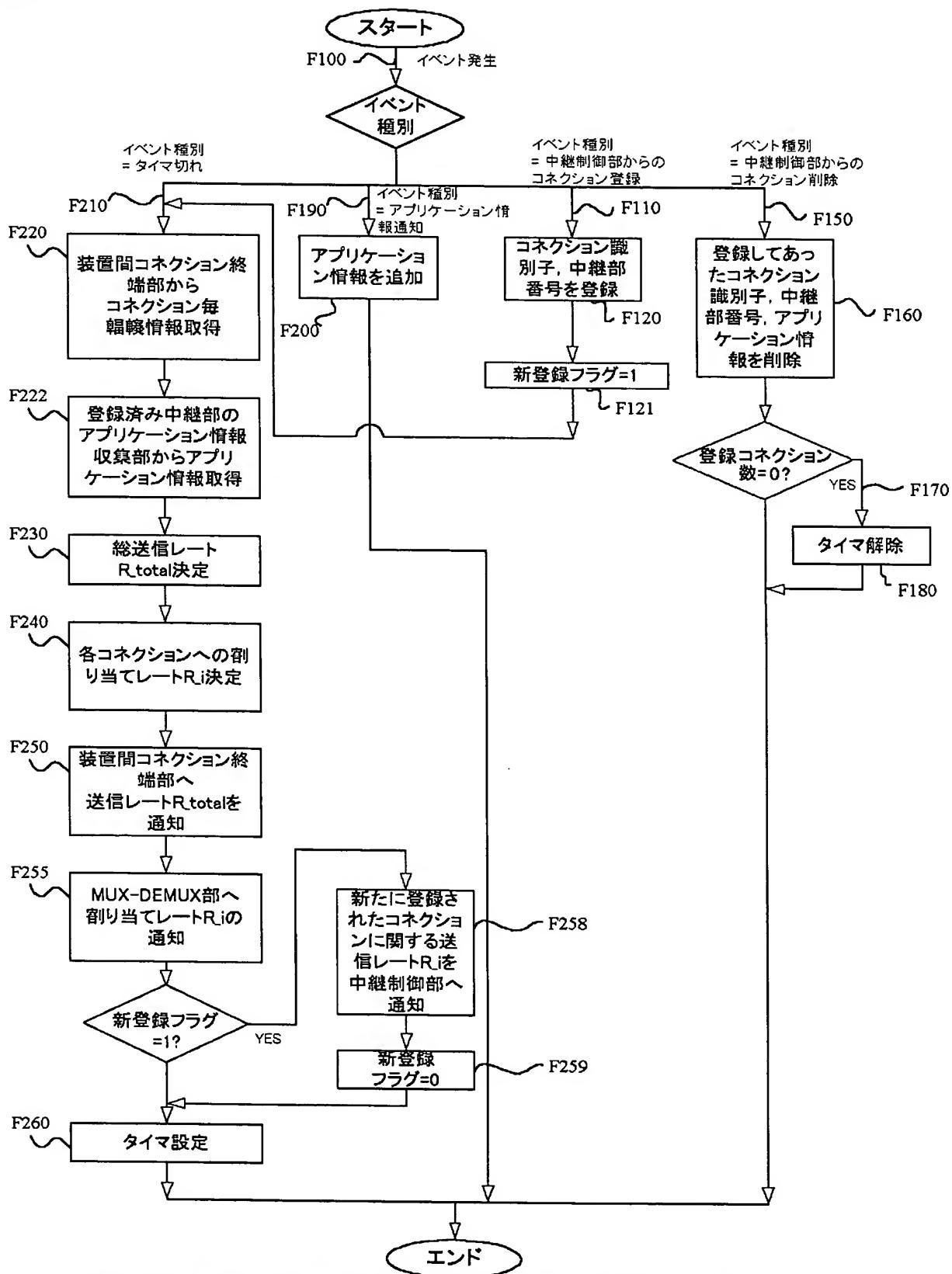
第7の実施の形態のトランスポート層中継装置を用いたネットワーク構成図

[図24]



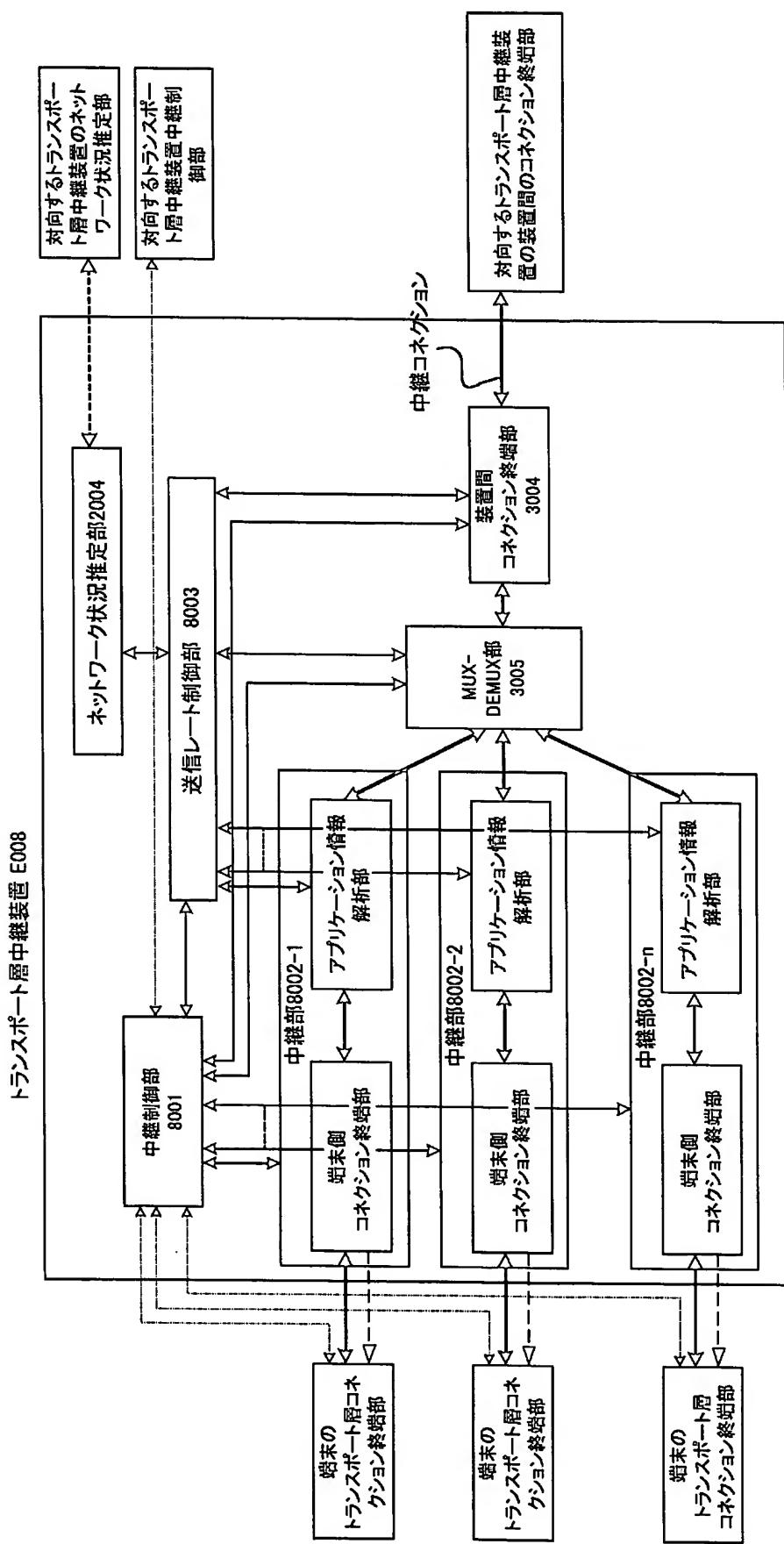
第7の実施の形態の動作説明タイムチャート

[図25]



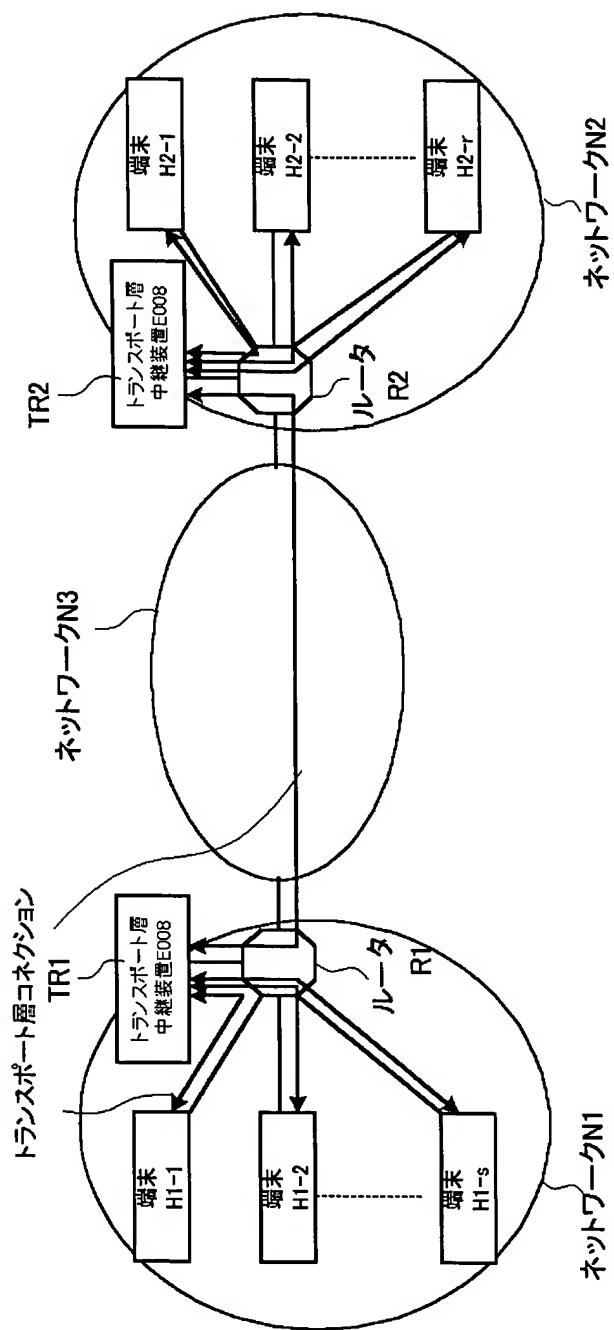
第7の実施の形態のレート制御部の動作フローチャート

[図26]



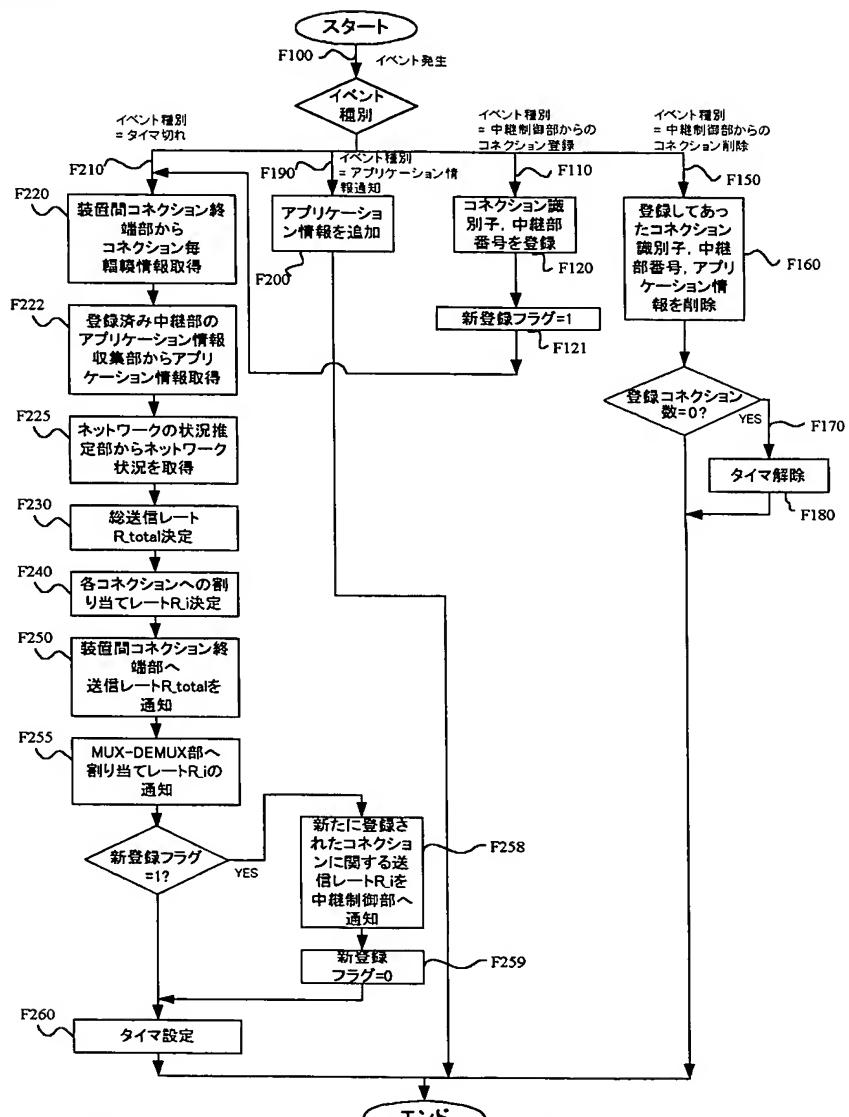
第8の実施の形態のトランスポート層中継装置の構成図

[図27]



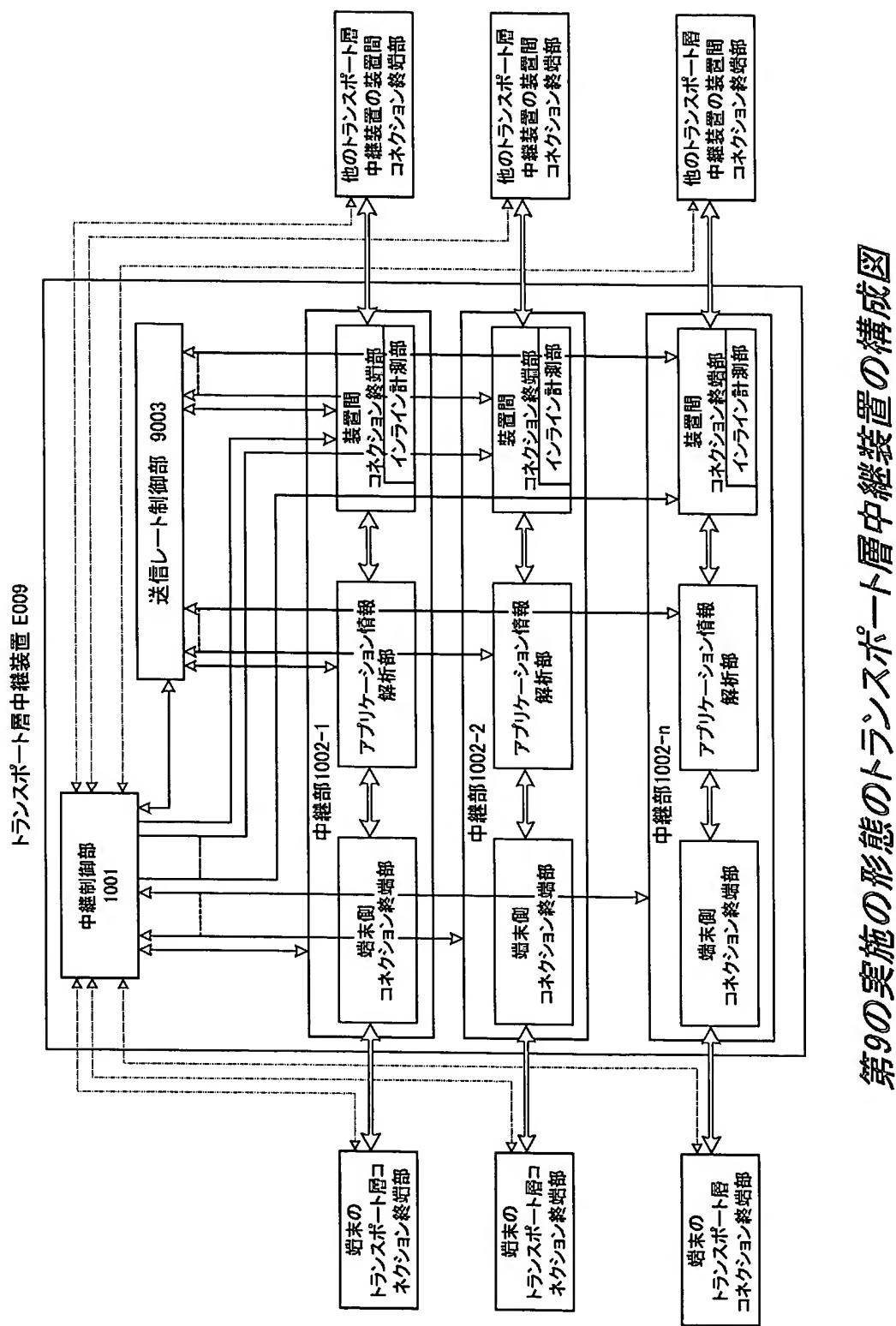
第8の実施の形態のトランスポート層中継装置を用いたネットワーク構成図

[図28]



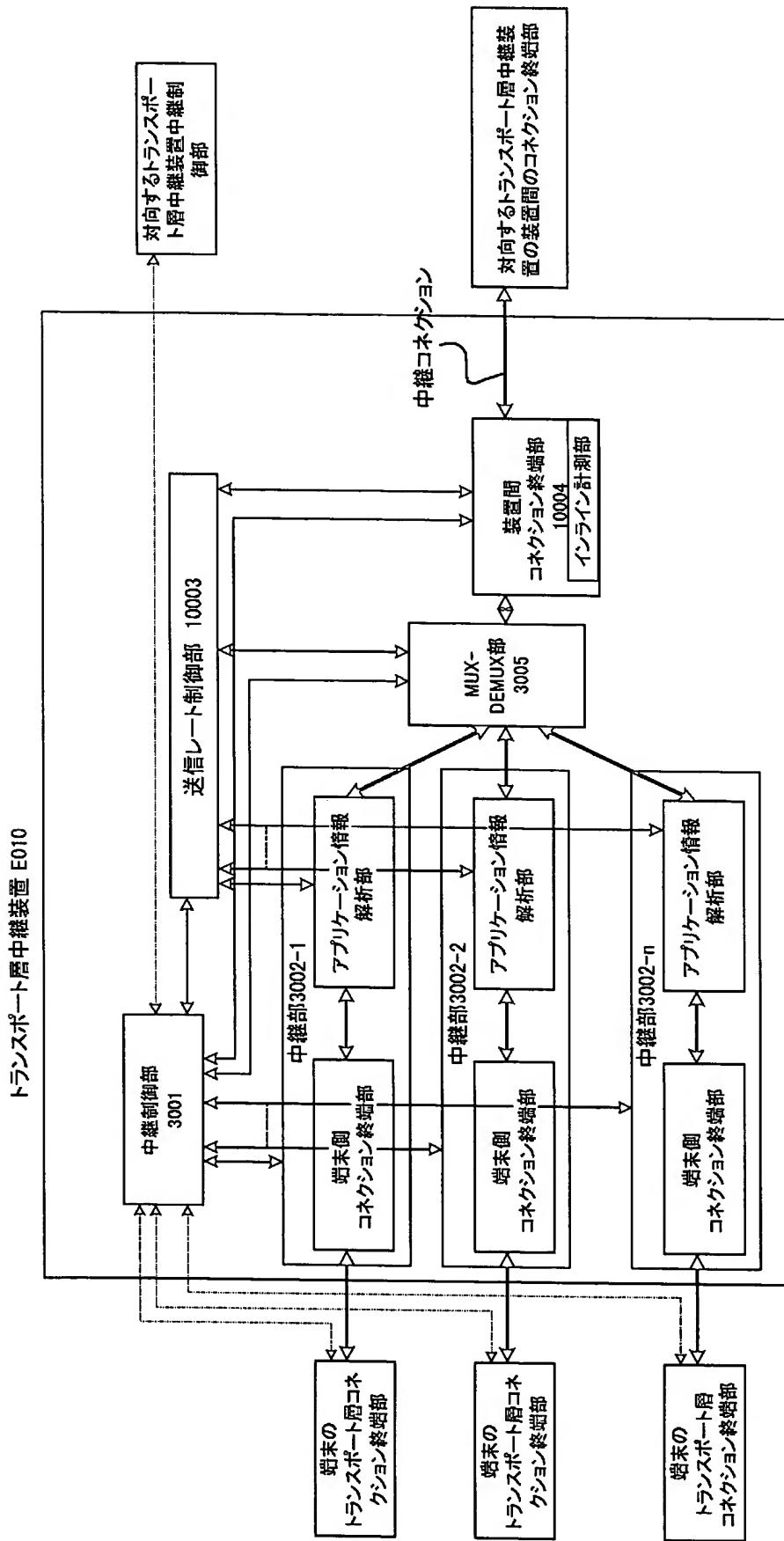
第8の実施の形態のレート制御部の動作フローチャート

[図29]



第9の実施の形態のトランスポート層中継装置の構成図

[図30]



第10の実施の形態のトランスポート層中継装置の構成図

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2004/009659

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl⁷ H04L12/46

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl⁷ H04L12/46

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

| | | | |
|---------------------------|-----------|----------------------------|-----------|
| Jitsuyo Shinan Koho | 1926-1996 | Toroku Jitsuyo Shinan Koho | 1994-2004 |
| Kokai Jitsuyo Shinan Koho | 1971-2004 | Jitsuyo Shinan Toroku Koho | 1996-2004 |

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|---|-----------------------|
| A | JP 2002-237841 A (NEC Corp.), 23 August, 2002 (23.08.02), Fig. 1 (Family: none) | 1-34 |
| A | JP 2003-198611 A (Hitachi, Ltd.), 11 July, 2003 (11.07.03), Fig. 1 (Family: none) | 1-34 |
| A | JP 2001-244968 A (Nippon Telegraph And Telephone Corp.), 07 September, 2001 (07.09.01), Fig. 1 (Family: none) | 1-34 |

 Further documents are listed in the continuation of Box C. See patent family annex.

| | |
|---|--|
| Special categories of cited documents: | |
| "A" document defining the general state of the art which is not considered to be of particular relevance | "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| "E" earlier application or patent but published on or after the international filing date | "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone |
| "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "O" document referring to an oral disclosure, use, exhibition or other means | "&" document member of the same patent family |
| "P" document published prior to the international filing date but later than the priority date claimed | |

Date of the actual completion of the international search
22 October, 2004 (22.10.04)Date of mailing of the international search report
09 November, 2004 (09.11.04)Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類 (国際特許分類 (IPC))
Int. C17 H04L 12/46

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))
Int. C17 H04L 12/46

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1926-1996年
日本国公開実用新案公報 1971-2004年
日本国登録実用新案公報 1994-2004年
日本国実用新案登録公報 1996-2004年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

| 引用文献の カテゴリー* | 引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示 | 関連する 請求の範囲の番号 |
|-----------------|--|------------------|
| A | J P 2002-237841 A (日本電気株式会社)、2002.08.23、図1 (ファミリー無し) | 1~34 |
| A | J P 2003-198611 A (株式会社日立製作所)、2003.07.11、図1 (ファミリー無し) | 1~34 |
| A | J P 2001-244968 A (日本電信電話株式会社)、2001.09.07、 図1 (ファミリー無し) | 1~34 |

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示す
もの
「E」国際出願日前の出願または特許であるが、国際出願日
以後に公表されたもの
「L」優先権主張に疑惑を提起する文献又は他の文献の発行
日若しくは他の特別な理由を確立するために引用する
文献 (理由を付す)
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって
出願と矛盾するものではなく、発明の原理又は理論
の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明
の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以
上の文献との、当業者にとつて自明である組合せに
よって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日
22. 10. 2004

国際調査報告の発送日

09.11.2004

国際調査機関の名称及びあて先
日本国特許庁 (ISA/JP)
郵便番号100-8915
東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)
石井 研一

5K 8124

電話番号 03-3581-1101 内線 3555